



B11

## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

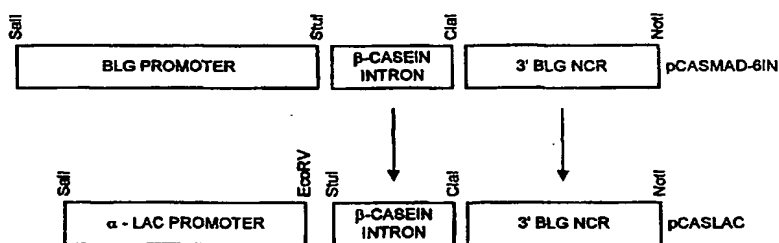
(51) International Patent Classification <sup>6</sup> : <b>C12N 15/00, 15/12, 15/85, 5/10, C07K 14/705, 14/78, 14/81, 16/00, A01K 67/027</b>		A1	(11) International Publication Number: <b>WO 99/03981</b>
			(43) International Publication Date: 28 January 1999 (28.01.99)
(21) International Application Number: PCT/GB98/02130		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 17 July 1998 (17.07.98)		<b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(30) Priority Data: 9715064.3 17 July 1997 (17.07.97) GB			
(71) Applicant (for all designated States except US): PPL THERAPEUTICS (SCOTLAND) LIMITED [GB/GB]; Roslin, Edinburgh EH25 9PP (GB).			
(72) Inventors; and (75) Inventors/Applicants (for US only): COLMAN, Alan [GB/GB]; PPL Therapeutics (Scotland) Limited, Roslin, Edinburgh EH25 9PP (GB). GARNER, Ian [GB/GB]; PPL Therapeutics (Scotland) Limited, Roslin, Edinburgh EH25 9PP (GB). DALRYMPLE, Michael, Alexander [GB/GB]; PPL Therapeutics (Scotland) Limited, Roslin, Edinburgh EH25 9PP (GB).			
(74) Agents: CORNISH, K., V., J. et al.; Kilburn & Strode, 20 Red Lion Street, London WC1R 4PJ (GB).			

(54) Title: HETEROLOGOUS EXPRESSION OF PROTEINS BY "RESCUED" VECTOR COMPRISING AN INTRON

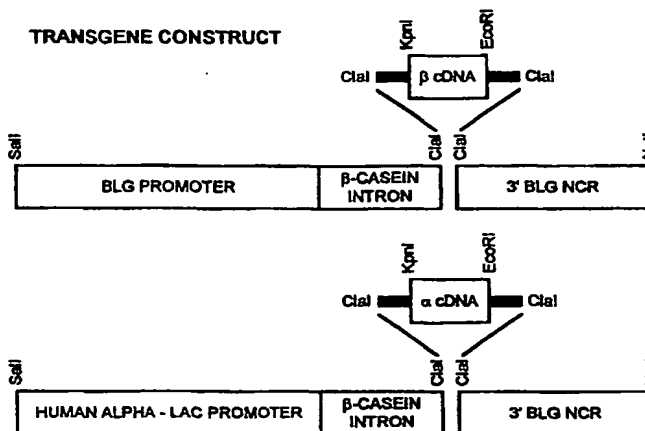
## (57) Abstract

A nucleic acid expression construct comprising: (a) a promoter; (b) an intron whose natural position is within the 5'-untranslated region of a gene from which it is derived; (c) a coding sequence; and (d) a 3'-flanking sequence wherein the intron (b) is not derived from the same gene as that from which either the promoter (a) or the protein-coding sequence (c) is derived and processes, vectors, hosts and uses involving such a construct to obtain *inter alia* an increase in the level of expression of the coding sequence.

## CLONING OF CASLAC



## TRANSGENE CONSTRUCT



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece	ML	Mali	TR	Turkey
BG	Bulgaria	HU	Hungary	MN	Mongolia	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MR	Mauritania	UA	Ukraine
BR	Brazil	IL	Israel	MW	Malawi	UG	Uganda
BY	Belarus	IS	Iceland	MX	Mexico	US	United States of America
CA	Canada	IT	Italy	NE	Niger	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NL	Netherlands	VN	Viet Nam
CG	Congo	KE	Kenya	NO	Norway	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NZ	New Zealand	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	PL	Poland		
CM	Cameroon	KR	Republic of Korea	PT	Portugal		
CN	China	KZ	Kazakhstan	RO	Romania		
CU	Cuba	LC	Saint Lucia	RU	Russian Federation		
CZ	Czech Republic	LI	Liechtenstein	SD	Sudan		
DE	Germany	LK	Sri Lanka	SE	Sweden		
DK	Denmark	LR	Liberia	SG	Singapore		
EE	Estonia						

## HETEROLOGOUS EXPRESSION OF PROTEINS BY "RESCUED" VECTOR COMPRISING AN INTRON

This invention relates to the expression of proteins in heterologous host systems, particularly in, but not limited to, the mammary gland of transgenic animals.

5

It has been shown, using regulatory DNA elements from milk protein genes, that it is possible to express heterologous proteins in the milk of transgenic livestock. One such gene, that for ovine  $\beta$ -lactoglobulin (BLG), has been cloned and characterised (Ali and Clark, *J. Mol. Biol.*, **199** 145-426(1988)). The authors subsequently demonstrated consistent, high level, expression of ovine BLG in the milk of mice transgenic for the entire gene (Simons *et al.*, *Nature*, **328** 530-532(1987); Harris *et al.*, *Developmental Genetics*, **12** 299-307(1991)). Further experiments demonstrated that the BLG promoter region can direct high levels of expression of a heterologous human protein to the milk of transgenic mice (Archibald *et al.*, *Proc. Natl. Acad. Sci. USA*, **87** 5178-5182(1990)). The generation of sheep, expressing human proteins in their milk using BLG regulatory elements, indicated that this technology was applicable to transgenic livestock (Simons *et al.*, *Bio/Technology*, **6** 179-183(1988); Clark *et al.*, *Bio/Technology*, **7** 487-492(1989)). The commercial feasibility of this technology, as a means of producing recombinant therapeutics in livestock milk, has been confirmed by the demonstration of high level expression of human  $\alpha_1$ -antitrypsin in the milk of transgenic sheep (Wright *et al.*, *Bio/Technology*, **9** 830-834(1991); Carver *et al.*, *Cytotechnology*, **9** 77-84(1992); Carver *et al.*, *Bio/Technology*, **11** 1263-1270(1993); Cooper and Dalrymple, *The Japanese Journal of Experimental Medicine, Developmental Biotechnology supplement*, **12(2)** 124-132(1994)).

This high level of expression of a heterologous protein in livestock milk was the result of using a fusion of the BLG promoter region to human genomic sequences (Wright *et al.*, *Bio/Technology*, **9** 830-834(1991)). Analogous cDNA based constructs were poorly expressed in transgenic mice (Whitelaw *et al.*, *Transgenic*

30

*Res.*, 1 3-13(1991)). Despite some notable exceptions in the field as a whole, (Ebert *et al.*, *Bio/Technology*, 9 835-838(1991); Velandar *et al.*, *Proc. Natl. Acad. Sci. USA*, 89 12003-12007(1992)) the general inefficient expression of cDNA based constructs is well documented (Brinster *et al.*, *Proc. Natl. Acad. Sci. USA*, 85 836-840(1988); Palmiter *et al.*, *Proc. Natl. Acad. Sci. USA* 88 478-482 (1991); Whitelaw *et al.*, *Biochem. J.*, 286: 31-39 (1992)). Observed problems include the influence of chromosomal position effects and distinct spatial and/or temporal expression in lines transgenic for the same construct. Such constructs can be improved by the addition of some natural or heterologous introns. However, expression levels from such constructs rarely match levels attained with constructs containing some or all natural introns in the region encoding a heterologous protein. The successful use of less than a full complement of introns is the subject of WO-A-9005188. In spite of that useful advance in the art, however, the genetic material encoding many potential target human proteins which may be produced by the transgenic mammary gland is very often, due to immediate non-availability or the size of the natural gene, limited to cDNAs. As such, a technique giving more consistent expression from transgene constructs containing intronless cDNA sequences is highly desirable.

A further advance in the expression of cDNAs is the so-called "rescue" technology, an approach developed by Clark and co-workers (Clark *et al.*, *Bio/Technology*, 10 1450-1454(1992); WO-A-9211358)) to overcome cDNA-related expression problems. It makes use of the observation that co-injection of an actively expressed transgene, such as the entire ovine BLG gene, together with an intronless construct results in the expression of the second construct where no expression is achieved when it is injected alone. Clark and colleagues have demonstrated the expression of up to 800 $\mu$ g/ml of human  $\alpha_1$ -antitrypsin (AAT) in the milk of mice transgenic for both BLG and an intronless human AAT construct. In mice transgenic for the latter construct alone, only one out of eight mice expressed and this at a level of only 3.9 $\mu$ g/ml. Similarly, using this technology,

an expression level of 108 $\mu$ g/ml from a wild type human protein C cDNA construct was achieved (WO-A-9211358). This represents approximately 20% the expression level obtained with an equivalent genomic based construct.

The cDNA construct alone gave no expression in 11 lines of transgenic mice.

5

The "rescue" phenomenon has been rationalised as follows. Strongly expressing genes have an innate ability to 'dominate' their chromosomal environment such that they are able to initiate and maintain a high expressing state. Intronless genes are deficient in some, as yet identified, feature which provides them with this capability. However, the dominant effect of the strong gene extends some way 5' and 3' to the gene itself and therefore by linking a 'weak' and 'strong' gene, some of the properties of the high expressing gene are conferred on the intronless gene. Clark and colleagues propose that this probably results in an open chromatin conformation associated with the actively expressing gene which encompasses adjacent intronless genes. The actively expressing gene may thus create a permissive domain allowing access to the intronless genes by the transcriptional machinery of the cell. In the absence of adjacent actively expressing genes, the intronless construct may be inaccessible, probably residing in condensed chromatin. Other possible explanations for this phenomenon include enhancer-like sequences present in the actively expressing gene but absent from the intronless construct interacting positively with the latter or simply that the actively expressing gene insulates the intronless gene from the negative effects of adjacent chromatin.

To take advantage of "rescue" technology, we have constructed a vector, pMAD, from the ovine BLG gene for the cloning of cDNAs (Figures 1 and 2). This vector contains the same 5' and 3' flanking sequences present in the BLG gene which itself always gives rise to high level expression in transgenic mice. However, it lacks all coding sequences and introns of the intact gene. Cloning of cDNAs in the unique *EcoRV* site between 5' and 3' flanking sequences results in constructs suitable for expression by the "rescue" approach. However, the issue of co-

30

injection of two covalently unlinked genes is not without its difficulties. There is always the risk that one gene or other is not represented in the final transgenic lines. Additionally, the two different genes may be present but not at the same locus. Subsequently they may segregate upon breeding. Finally, the physical  
5 structure of a BLG/pMAD array is not determined prior to injection and there is no control over it. The relative copy numbers of the two genes may vary especially if the DNA concentrations of the two constructs are not tightly controlled.

cDNAs have been successfully expressed at high levels, in a limited number of  
10 cases. It is not clear from the literature why this should be the case. However, the fact is that a cDNA has never (to our knowledge) been expressed at high levels from a BLG construct other than by rescue.

We had noted the work of Brinster and Palmiter (*ibid*) and others and we sought to  
15 incorporate a BLG intron into our cDNA constructs. To this end the vector pMAD6 was constructed, containing almost all the BLG sequence 3' to the natural BLG stop codon, i.e. a portion of exon 6, intron 6, all of exon 7 and those available sequences downstream of the polyadenylation site (see Figures 1 and 2). A protein C cDNA in this vector (pCORP3) expresses at detectable levels but not  
20 nearly as well as "rescued" intronless pCORP2 (see table 2). Thus we can conclude that the mere presence of a BLG intron is insufficient to achieve high level expression.

Noting that certain genes have an intron in the 5'-untranslated region (5'-UTR), we  
25 engineered the natural BLG first intron into the 5'-UTR of the BLG sequences in pMAD (to give pMAD1) and into pMAD6 (to give pMAD16). When protein C cDNAs were put into these vectors, there was no detectable expression of protein C in the milk of lactating female transgenics (see table 2; pCORP6). This indicates that the mere presence of intronic sequences in the 5'-UTR of a gene is in general  
30 insufficient to allow expression of a cDNA.

We have now found, however, that if, instead of the BLG first intron, an intron whose natural position is within the 5'-untranslated region of its gene is used, good expression results.

5 According to a first aspect of the invention, there is provided an expression construct comprising:

- (a) a promoter;
- 10 (b) an intron whose natural position is within the 5'-untranslated region of a gene from which it is derived;
- (c) a coding sequence; and
- 15 (d) a 3'-flanking sequence,

wherein the intron (b) is not derived from the same gene as that from which either the promoter (a) or the coding sequence (c) is derived, and, in particular wherein the promoter (a) drives expression of the coding sequence (c) at a level which is  
20 elevated by virtue of the presence of the intron (b).

Elevated levels of expression include expression where previously none was measurable (or obtained). Elevated levels is optionally defined as a level higher than obtained by the construct without the intron (and optionally the 3' flanking  
25 sequence) described above.

Preferably the expression construct is a DNA expression construct. Preferably the coding sequence is a protein-coding sequence although it may code non-protein substances such as ribozymes. The construct is effective for two particular  
30 reasons; firstly, the promoter (a) drives expression of the coding sequence (c) at a

level which is elevated by virtue of the presence of the intron (b) and/or secondly the coding sequence (c) is more likely to be expressed in a transgenic host, by virtue of the presence of the intron (b). The second effect is particularly important when taking into account the length of time and the efforts required to produce transgenic animals useful as bioreactors for the production of useful proteins, etc. It is also important in laboratory scale trials to determine and obtain transgenic hosts. Use of constructs as described in the claims have shown that an increased number of transgenic hosts express the coding sequence over use of the constructs without specific intron described herein (e.g see number of expressing founders in Table 3). The elevated level of expression of the coding sequence and/or the expression of the coding sequence may be by virtue of the presence of the intron (b) and the 3'-flanking sequence (d).

The DNA expression construct may be useful for expression in any suitable host system such as, for example, prokaryotes, (e.g. *E.coli*), fungi, plant and animal (including mammalian) cell lines and transgenic plants and animals (including mammals). However, it is in transgenic animal hosts that the expression constructs of the invention are most useful. In principle, the invention is applicable to all animals, including birds such as domestic fowl, amphibian species and fish species. The protein may be harvested from body fluids (such as milk, blood or urine) or other body products (such as eggs, where appropriate). In practice, it will be to (non-human) mammals, particularly placental mammals, that the greatest commercially useful applicability is presently envisaged. This is because expression in the mammary gland, with subsequent optional recovery of the expression product from the milk, is a proven and preferred technology. It is with ungulates, particularly economically important ungulates such as cattle; sheep, goats, water buffalo, camels and pigs that the invention is likely to be most useful. The generation and usefulness of such mammalian transgenic mammary expression systems is both generally, and in certain instances specifically, disclosed in WO-A-8800239 and WO-A-9005188.



In this text, the meaning of a sequence being derived from a gene does not require that the sequence has actually been obtained from the gene in question. Rather, all and any copy, as well as the original sequence is meant. Further, any modification  
5 to the sequence which does not remove the desired end result can be used.

In addition to being useful in transgenic animal expression for non-therapeutic purposes (as far as the host is concerned), constructs of the invention may also be useful in genetic therapy in humans or other animals.

10

The promoter can be any suitable promoter chosen from a gene different from the source of the intron (b). Within that constraint, it will be chosen having regard to its desired properties in the construct of the expression system to be used and its ability to derive expression of heterologous sequences in cell culture or in a  
15 transgenic organism. A promoter is any sequence which drives expression of a coding sequence. For example, the BLG promoter does not express particularly highly in cells which do not respond to prolactin (such as COS cells). A 'cell' promoter according to the invention is the HCMV (human cytomegalovirus) IE gene promoter. Other promoters of the invention include, endothelial promoters  
20 such as vascular cell adhesion molecule (VCAM), platelet endothelial cell adhesion molecule-1 (PECAM), inter-cellular adhesion molecule-2 (ICAM) and smooth muscle promoters, such as Desmin E and Desmin P. A preferred promoter is one which drives expression of the protein coding sequence in mammalian cells. In relation to expression in transgenic animal hosts, the preferred expression system  
25 involves expression in the mammary gland of transgenic placental mammals. For this purpose, milk protein promoters will generally be used, preferably but not necessarily derived from the species chosen as an expression host. The promoter may be a casein promoter (such as an  $\alpha$ -,  $\beta$ - or  $\kappa$ -casein promoter), but it is preferred that it be a non-casein promoter, such as the human Bile Salt Stimulated  
30 Lipase (BSSL) promoter, more preferably a whey protein promoter, such as that of

whey acidic protein (WAP),  $\alpha$ -lactalbumin or, most preferred of all,  $\beta$ -lactoglobulin. Figure 3 is a schematic representation of the cloning of pCASLAC and obtaining transgene constructs therefrom. pCASLAC corresponds to pCASMAD6 (see Fig. 2, 4 and 7) only using the more tightly regulated  $\alpha$ -lac promoter. Of course, the present invention covers promoters, in the constructs described, which have not yet been isolated or characterized. One general way for isolating specific promoters (such as mammalian promoters) for use in the present invention is to isolate specific cDNAs by differential display or from subtractive cDNA libraries. These, in turn, are used to screen genomic libraries for the cognate promoters.

In addition, the present invention encompasses the use of a modified low expressing naturally occurring promoter *in vitro* to an increased level of expression (eg. by addition of an enhancer) or to use a promoter with a higher level of expression in a crossed species (eg. the human  $\alpha$ -lactalbumin promoter expresses better in mice than the endogenous mouse promoter).

A promoter according to the invention may also be a viral or modified cellular promoter or a completely artificial promoter having the properties of high level expression (preferably mammalian species). Details of suitable promoters can be found in Houdibine, J-M., *J. Biotech.*, 34: 269-287 (1994); Garner, I. & Dalrymple, M., in *"Encyclopedia of Molecular Biology: Fundamentals and Applications"*, Robert A. Myers (Ed.), Weinheim, NY.

Element (b) of a construct of the invention is an intron whose natural position is within the 5'-untranslated region (5'-UTR) of its natural gene (i.e the gene with which it is naturally associated). The whole intron is not necessarily required. Fragments or portions may be sufficient. The requirement for the present invention is that the level of protein expression, from any construct according to the invention, is elevated by virtue of the presence of the intron, or parts thereof. It

has been shown that the first and third portions of an intron (which has been divided into three fairly equal parts), recombined, are often effective. Generally speaking, the intron for inclusion in a construct according to the invention will be the first intron of such a gene. Examples of genes with such known introns include; human and rat aldolase A, human type II IL-1 receptor, human UDP-N-acetylglucosaminyl transferase, mouse involucrin and mouse adenosine deaminase. Some genes have more than one intron whose natural positions are all within the 5' untranslated region of its natural gene. The present invention recognises this and covers, within element (b), one or more of such introns (for example in a gene with two introns naturally positioned in the 5' UTR, they may separately, together or parts of each cojoined be included in a part of a construct according to the invention). These and other yet unidentified introns whose natural position is within the 5' untranslated region of its natural gene may be used according to the present invention. Also included are: the introns of several gene families including; the actin family (two skeletal muscle actins-alpha cardiac and alpha skeletal, two smooth muscle actins-alpha smooth and gamma smooth, and two non-muscle actins-beta and gamma cytoplasmic actin), the troponin family (cardiac, skeletal and foetal troponins) and the casein family ( $\alpha$  S1,  $\alpha$  S2,  $\beta$  and  $\kappa$ ). Preferably the intron is the first intron of the family. In the case of transgenic mammary specific expression, the most preferred gene family from which the intron may come is the casein family or the actin family. The intron may preferably be from the same source of organism as the promoter and/or the expression system which it is in (e.g. mammalian, bovine, ovine, etc.).

DNA expression constructs of the present invention are different from that of Barash *et al.* (*Nucl. Acids Res.* 24(4) 602-610 (1996)), in that Barash *et al.*'s constructs include  $\beta$ -lactoglobulin intragenic sequences which are not within the 5'-untranslated region. Barash *et al.* do not refer to the possibility of using an intron whose natural position is within the 5'-untranslated region of its natural gene.

Caseins, whose genes represent a preferred source of the 5'-UTR introns useful in

the present invention, are the major mammalian milk proteins and are encoded by a small gene family, which in cows and sheep consists of four members,  $\alpha_{s1}$ ,  $\beta$ ,  $\alpha_{s2}$  and  $\kappa$ , and in mice and rats five,  $\alpha$ ,  $\beta$ ,  $\gamma$ ,  $\epsilon$  and  $\kappa$  (Yu-Lee & Rosen, *J. Biol. Chem.*, **258** 10794-10804 (1983); Jones *et al.*, *J. Biol. Chem.*, **206** 7042-7048(1985); Thompson *et al.*, *DNA*, **4** 263-271 (1985)/ reviewed by Mercier & Vilotte, *J. Dairy Sci.*, **76** 3079-3098(1993)). The evolution of the calcium sensitive caseins ( $\alpha$  and  $\beta$ ) is believed to have occurred by recruitment of exons encoding discrete functional domains, followed by intragenic and intergenic duplication to create the present number of similar exons within a given gene, and of genes within a family (Jones *et al.*, *J. Biol. Chem.*, **206** 7042-7048(1985); Groenen *et al.*, *Gene*, **123** 187(1993); reviewed by Mercier & Vilotte, *J. Dairy Sci.*, **76** 3079-3098(1993)). There is no evidence that  $\kappa$  casein is evolutionally related to the other caseins. Both in sequence homology and protein function it appears to be related to  $\gamma$  fibrinogen (Jolles *et al.*, *Biochim. Biophys. Acta.*, **365** 335(1974); Thompson *et al.*, *DNA*, **4** 263-271(1985); Alexander *et al.*, *Eur. J. Biochem*, **178** 395-401(1988)), which performs a cleavage-induced clotting function in blood similar to the clotting function of  $\kappa$  casein in the stomach. The caseins all map to a single chromosome in rodents, sheep, cows, humans and pigs (reviewed by Mercier & Villotte, *J. Dairy Sci.*, **76** 3079-3098(1993)), all four bovine caseins have been mapped to a single 250 Kb locus (Ferretti *et al.*, *Nucleic Acids Res.*, **18** 6829-6833(1990); Threadgill & Womack, *Nucleic Acids Res.*, **18** 6935-6942(1990)) and all five mouse caseins to a 400 Kbp region (Tomlinson *et al.*, *Mammalian Genome*, **7** 542-544).

The first intron of the calcium sensitive casein genes is naturally positioned in the 5'-UTR, upstream of the start of translation. The position of this intron is conserved across species barriers, indicating that there may be some critical function for an intron in this position. The intron may be obtained by PCR amplification from genomic DNA. The resulting DNA fragment may be cloned into a suitable site of an appropriate vector, such as the pMAD6 vector described

above.

Constructs of the invention also contain a coding sequence (c), whose expression is driven by the promoter (a) under the beneficial influence of the intron (b). The protein-coding sequence may code for any (natural or modified) protein of interest, particularly those which may be advantageously produced in the preferred mammary gland expression systems. Examples of classes of such proteins, and specific instances within those classes, are as follows: blood proteins involved in haemostasis including factors V, VII, VIII, IX, X, XIII, PAI-1, PAI-2, TFPI, protein C (details of protein C according to the present invention can be found for example in EP-A-191606 and W097/20043), protein S, alpha 1-antitrypsin (AAT) (details of which can be found in general from Perlino *et al.* EMBO Journal, 6, 2767-2771, 1987 and W090/05188), tPA, fibrinogen (details for which may be found in W095/23868 and the references cited therein); other protease inhibitors such as serpins, Kazal/Kunitz inhibitors, kinninogens, stefins, cystatins or tissue inhibitors of metalloproteinases; growth factors; protein hormones; structural proteins such as collagens (details of which may be found in W093/07889, W094/16570, W097/08311 and the references cited in these publications) and keratins; enzymes such as lipases, other proteases and transferases; and antibodies.

While the protein-coding sequence may in principle be any suitable sequence, such as either the full natural genomic structure, a minigene sequence consisting of some, but not all, of the introns naturally present in the gene, or a cDNA (containing no introns), it will generally be with cDNA sequences that the invention is most useful. This is because the invention may conveniently enable the expression of protein from cDNAs which may otherwise only be achievable using minigenes or full genomic sequences. Furthermore, some proteins may be expressed in nature from intronless genes (e.g. bacterial or yeast genes, human thrombomodulin) or have natural intron structures incompatible with the chosen host (e.g. invertebrate or plant genes in a mammalian cell). In these cases the 'cDNA' route is the only one available.

The intron (b) is preferably positioned upstream of the translation start site for the protein-coding sequence (c), by analogy with its position in its natural environment.

5

Particularly preferred constructs according to the present invention are the BLG promoter with either (i) the first intron from bovine  $\beta$ -casein or (ii) the first intron from muscle cardiac actin or (iii) the first intron from ovine  $\beta$ -casein. More preferably, the 3' flanking sequence is from BLG as described below for preferred  
10 3' sequences under (i). Particularly preferred constructs of the present invention include the following:

- (i) BLG promoter + bovine  $\beta$ -casein intron 1 + BLG 3' sequence (particularly the 3' sequence beginning immediately 3' to the natural  $\beta$ -lactoglobulin stop codon and continuing to at least about 30 bases 3' of the poly-A site), optionally including ovine beta-lactoglobulin intron 6 (preferred positioned 5' to the flanking sequence and 3' to any coding sequence); in particular the construct pCASMAD6 as described in Fig. 2, 4 or 7;  
15
- (ii) BLG promoter + muscle cardiac actin intron 1 + BLG 3' sequence (particularly the 3' sequence beginning immediately 3' to the natural  $\beta$ -lactoglobulin stop codon and continuing to at least about 30 bases 3' of the poly-A site), optionally including ovine beta-lactoglobulin intron 6 (preferred position 5' to the flanking sequence and 3' to any coding sequence); in  
20 25 particular the construct pACTMAD6 as described in Fig. 2, 5 or 7;
- (iii) BLG promoter + ovine  $\beta$ -casein intron 1 + ovine  $\beta$ -casein 3' flanking sequence; in particular the construct pBOB as described in Fig 2, 6 or 7.  
30

Preferably, following the coding sequence will be such 3'-sequences as may be necessary or appropriate. In the invention at its broadest, it is not thought that the nature of such 3'-sequences is particularly limited. The 3'-flanking sequence may or may not include its natural intron. Suitable 3' flanking sequences preferably  
5 comprise functional elements which are able to direct the correct transcription, termination and 3' end processing. These can be determined, without undue burden, by the person skilled in the art. However, certain 3'-flanking sequences have been found to be particularly useful. These include, but are not restricted, to:  
(i) a poly-A site (poly A addition site), (ii) a  $\beta$ -lactoglobulin gene 3'-sequence  
10 beginning immediately 3' to the natural  $\beta$ -lactoglobulin stop codon and continuing to at least about 30 bases 3' of the poly-A site (as found in pMAD6 and pCASMAD6 and PACTMAD6), or (iii)  $\beta$ -casein 3' sequences including poly A signal. These sequences (as used in pBOB) consist of 6.5Kbp of DNA incorporating ovine  $\beta$ -casein exons 7 to 9, introns 7 and 8, and approximately  
15 4.8Kbp of 3' sequence.

The presence of such 3'-sequences in the construct adds stability to it. It is believed that the relative orientation of the first and last intron may contribute to this stability.

20

The  $\beta$ -lactoglobulin gene 3'-sequence may be cloned from a  $\beta$ -lactoglobulin gene or amplified by PCR and cloned from genomic DNA, which may be of ovine origin. As mentioned above, it begins with the natural  $\beta$ -lactoglobulin gene sequence immediately 3' to the stop codon, which is a TAG codon occurring in  
25 exon VI. It extends to at least about 30 bases 3' of the poly-A site, which is in exon VII. Exons VI and VII bracket intron 6 which is present in its entirety. The preferred minimum length of the  $\beta$ -lactoglobulin-derived 3'-sequences is about 2.3 Kb. For additional preference at least about 50 bases 3' to the poly A site are present. Similarly, the  $\beta$ -casein 3' sequences may be cloned from the  $\beta$ -casein  
30 gene or amplified by PCR and cloned from genomic DNA, which may be of ovine

origin.

Appropriate signal and/or secretory sequences, operably linked to the construct may be present if necessary or desirable.

5

In other aspects, the invention is directed to:

- 10 • a process for the preparation of a construct according to any feature of the first aspect. The process comprises linking together selected nucleotide bases and/or nucleotide sequences;
- 15 • a vector comprising a construct according to the first aspect of the invention. The vector may be plasmid, phage, cosmid or other vector type, for example derived from yeast. The vector may be an expression vector;
- a process for the preparation of a vector described above, comprising introduction of a construct according to the first aspect of the invention into a vector construct;
- 20 • a process for the preparation of a host (preferably an expression host), the process comprising introducing a DNA expression construct (or a vector), as described above, into a suitable organism; the process, in particular provides a host which expresses elevated levels of the coding sequence in the construct;
- 25 • a host organism (preferably an expression host organism) incorporating a DNA expression construct (or a vector) as described above (and preferably capable of giving rise to expression of protein encoded by the construct, although non-expressing hosts such as *Escherichia coli* and other
- 30 procaryotes may be useful as cloning hosts); The host may be a eukaryotic



or prokaryotic cell/organism, such as bacteria, insect or yeast cells, as well as animal tissues (cells in culture) and animals themselves. Such animals are transgenic and preferred transgenic animals include mammals, in particular non-human placental mammals such as pigs, sheep, cattle and goats. Preferably the host (e.g. transgenic animal) according to the invention has the construct (of the first aspect of the invention) integrated into its genome. It is particularly preferred that the transgenic animal transmits the construct to its progeny, thereby enabling the production of at least one subsequent generation of producer animals. Such a host organism, in particular expresses elevated levels of the coding sequence in the construct;

- a process of preparing a protein, the process comprising allowing an expression host to express a DNA expression construct as described above, and optionally subsequently purifying the protein;
- a protein when prepared by such a process. The protein may be a fusion protein;
- the use of a nucleic acid expression construct comprising a promoter, an intron whose natural position is within the 5'-untranslated region of a gene from which it is derived, a coding sequence and a 3' flanking sequence to obtain a transgenic host, preferably with elevated levels of the expressed coding sequence;
- the use of a nucleic acid construct comprising a promoter, an intron whose natural position is within the 5' untranslated region of a gene from which it is derived, a coding sequence and a 3' flanking sequence to increase the likelihood of expression of the coding sequence from a transgenic host which incorporates the nucleic acid construct;

- a process for improving whether an individual or a number of transgenic hosts express a transgene coding sequence, the process comprising introducing into a host, a nucleic acid construct comprising a promoter, an intron whose natural position is within the 5' untranslated region of a gene  
5 from which it is derived, a coding sequence and a 3' flanking sequence.

In addition to the construct according to the first aspect of the invention, there is provided an empty 'cassette', including all features in claim 1, without the coding sequence. Such a "cassette" provides an easy means by which to provide a high  
10 expressing vector for other parties to use by simply introducing coding sequences of interest by restriction endonuclease cutting of the empty cassette and religation (according to standard techniques). The empty "cassette" is for use with an incorporated coding sequence of interest.

15 Preferred features for each aspect of the invention are as for each other aspect *mutatis mutandis*.

The present invention also provides, as a separate aspect, the novel expression of collagen cDNA (natural procollagen chains or modified collagen). Preferably the  
20 collagen cDNA is expressed via a construct according to the first aspect of the invention. Preferred features of and all different aspects of the invention described herein above in relation to the construct, also apply to the expression of collagen cDNA. Particular preferred details in relation to collagen are described above under a discussion of the protein-coding sequences, including references thereto.

25 For example, for the expression of all collagen DNA (cDNA or otherwise), expression hosts may co-express prolyl 4-hydroxylase, which is a post-translational enzyme important in the natural biosynthesis of procollagen.

The invention will now be illustrated by the following examples. The examples  
30 refer to the drawings, in which:

FIGURE 1 shows the  $\beta$ -lactoglobulin (BLG) sequences and the plasmids pMAD and pMAD6.

FIGURE 2 shows the origin of sequences present in plasmids pMAD, pMAD6, pCASMAD6 and pACTMAD6.

FIGURE 3 shows the construction of pCASLAC

FIGURE 4 shows the construction of pCASMAD6.

FIGURE 5 shows the construction of pACTMAD6.

FIGURE 6 shows the construction of pBOB.

FIGURE 7 shows details of pMAD, pMAD6, pCASMAD6, pACTMAD6 and pBOB.

Preferred embodiments of the invention are based on the use of the BLG promoter, and are designed to express cDNAs from the BLG gene. The structure of pMAD6 is indicated in Figures 1, 2 and 7. This vector contains the same 5' and 3' flanking sequences present in the ovine BLG gene which itself always gives rise to high level expression in transgenic mice.

However, it lacks all protein coding sequences and introns 1 to 5 of the intact gene. The 3' non coding exons of the gene remain in this vector together with the final intron of the BLG gene. Cloning of cDNAs in the unique *EcoRV* site between 5' and 3' flanking sequences results in constructs suitable for expression of cDNAs. Incorporation of the BLG 3' sequences are not essential for the invention. Such BLG 3' sequences can be substituted by any competent 3' flanking sequences with or without an intron situated downstream of the last (stop) codon of such a gene.

In outline, the first intron of the bovine  $\beta$ -casein gene was amplified by PCR from genomic DNA. The resulting DNA fragment, of approximately 2 Kbp, was cloned and subsequently subcloned into the *EcoRV* site of pMAD6 in such a way that the  
5 original *EcoRV* site was destroyed and reformed on the 3' side of the intron. This gave the vector pCASMAD6 (Figures 1, 2, 4 and 7). A cDNA encoding human protein C was inserted into the unique *EcoRV* site of pCASMAD6 and the new construct called pCORI69. The cDNA utilised encodes a mutant form of the natural protein C (PC962): the mutation was designed to allow more efficient  
10 processing of the mature protein (Foster *et al*, *Biochemistry*, 29:347-354 (1990)).

This mutant form of the human protein C cDNA has been incorporated into a construct pCORP9, exactly analogous to pCORP2 (see WO-A-9211358). In "rescue" experiments pCORP9 expressed particularly poorly, the highest  
15 expressing line being 3 $\mu$ g/ml compared to 108 $\mu$ g/ml for the wild type cDNA. This indicates that this mutant cDNA is particularly difficult to express at high levels and therefore is a very exacting test of any cDNA expression system.

All references to the DNA sequence of the  $\beta$ -lactoglobulin gene utilise the numbering of the sequence allocated EMBL Accession No. X12817 (Harris *et al.*,  
20 NAR 16: 10379-80 91988).

## EXAMPLES

### General

25

Where not specifically detailed, recombinant DNA and molecular biological procedures were after Maniatis *et al* ("Molecular Cloning" Cold Spring Harbor (1982)) "Recombinant DNA" Methods in Enzymology Volume 68, (edited by R. Wu), Academic Press (1979); "Recombinant DNA part B" Methods in  
30 Enzymology Volume 100, (Wu, Grossman and Moldgave, Eds), Academic Press

- (1983); "Recombinant DNA part C" Methods in Enzymology Volume 101, (Wu, Grossman and Moldgave, Eds), Academic Press (1983); and "Guide to Molecular Cloning Techniques", Methods in Enzymology Volume 152 (edited by S.L. Berger & A.R. Kimmel), Academic Press (1987). Unless specifically stated, all chemicals
- 5 were purchased from BDH Chemicals Ltd, Poole, Dorset, England or the Sigma Chemical Company, Poole, Dorset, England. Unless specifically stated all DNA modifying enzymes and restriction endonucleases were purchased from BCL, Boehringer Mannheim House, Bell Lane, Lewes, East Sussex BN7 1LG, UK.
- 10 [Abbreviations: bp = base pairs; kb - Kilobase pairs, AAT =alpha1-antitrypsin; BLG = beta-lactoglobulin; FIX = factor IX; *E. coli* = Escherichia coli; dNTPs = deoxyribonucleotide triphosphates; restriction enonucleases are abbreviated thus e.g. BamHI; the addition of -O after a site for a restriction endonuclease e.g. PvuII-O indicates that the recognition site has been destroyed].

15

### Construction of Plasmids

#### Vectors

#### 20 Plasmid pUCPM

- The multiple cloning site of the vector pUC18 (Yanisch-Perron *et al.*, (1985) Gene 33: 103-119) was removed and replaced with a synthetic, double stranded, oligonucleotide containing the new restriction sites: *Pvu*I/*Mlu*I/*Sal*I/*Eco*RV/*Xba*I/*Pvu*I/ *Mlu*I, and flanked by 5'-overhangs compatible with the restriction sites
- 25 *Eco*RI and *Hind*III. pUC18 DNA was cleaved with both *Eco*RI and *Hind*III and the new linker DNA was ligated into pUC18. The DNA sequence across the new multiple cloning site was confirmed. This new vector was called pUCPM.

#### Plasmid pUCXS

- 30 The  $\beta$ -lactoglobulin gene sequences from plasmid pSS1tgXS (see WO-A-9201358)

were excised on a *SalI/XbaI* fragment and recloned into the vector pUCPM, cut with *SalI* and *XbaI*, to give plasmid pUCXS.

#### Plasmid pUCXS/RV

5 The plasmid pSSltgSE (see WO-A-8800239) contains:  $\beta$ -lactoglobulin gene sequences from the *SphI* site at position 754 to the *EcoRI* site at 2050, a region spanning a unique *NotI* site at position 1148. This insert contains a single *PvuII* site (832) which lies in the 5'-untranslated region of the  $\beta$ -lactoglobulin mRNA. Into this site was blunt-end ligated a double stranded, 8bp, DNA linker encoding  
10 the recognition site for the enzyme *EcoRV*, to give the plasmid pSSltgSE/RV. The DNA sequences bounded by *SphI* and *NotI* were then excised and used to replace the equivalent fragment in the plasmid pUCXS, thus effectively introducing a unique *EcoRV* site into the  $\beta$ -lactoglobulin gene placed in such a way as to allow the insertion of any additional DNA sequences under the control of the  $\beta$ -  
15 lactoglobulin gene promoter and 3' to the initiation of transcription. The resulting plasmid was called pUCXS/RV.

#### Plasmid pUCSV

A derivative of pUCXS/RV, containing only the 4.3 Kbp of the  $\beta$ -lactoglobulin gene which lie 5' to the transcription initiation site (the promoter), was constructed  
20 by subcloning the *SalI-EcoRV* fragment into pUCPM; this plasmid is called pUCSV.

#### Plasmid pBLAC100

25 A fragment of the 3' flanking sequence of the  $\beta$ -lactoglobulin gene was subcloned in such a way as to eliminate all introns. Plasmid DNA of pUCXS/RV was partially digested with *SmaI* by performing an enzyme titration with lower and lower concentrations of enzyme at a fixed DNA concentration. The *SmaI* protein was removed by phenol-chloroform extraction and ethanol precipitation and the  
30 DNA resuspended in water. This DNA was subsequently digested to completion

with the enzyme *Xba*I. DNA cut once at the *Sma*I site, position 5286 and then cleaved with *Xba*I gave a characteristic band of size 2.1 Kbp. This band was purified from an agarose gel slice and ligated into *Sma*I and *Xba*I cut pBSIISK+ (Stratagene Ltd., Cambridge Science Park, Cambridge, UK) to give the plasmid pBLAC100.

#### Plasmid pMAD

The  $\beta$ -lactoglobulin cloning vector pMAD was constructed to allow rapid insertion of cDNAs under the control of the  $\beta$ -lactoglobulin gene promoter and 3'-flanking sequences. Such constructs contain no introns. The plasmid pBLAC100 was opened by digestion with both *Eco*RV and *Sal*I, the vector fragment was gel purified. Into this was ligated the 4.3 Kbp promoter fragment from the plasmid pUCSV as a *Sal*I-*Eco*RV fragment. This construct is termed pST1 and constitutes a  $\beta$ -lactoglobulin mini-gene encoding the 4.3 Kbp promoter and 2.1 Kbp of 3' flanking sequences. A unique *Eco*RV site is present to allow blunt-end cloning of any additional DNA sequences. In order to allow excision of novel  $\beta$ -lactoglobulin gene constructs with the enzyme *Mlu*I the entire mini-gene from pST1 was excised on a *Xho*I-*Not*I fragment, the DNA termini made flush with Klenow polymerase, under standard conditions, and blunt-end cloned into the *Eco*RV site of pUCPM to give pMAD.

#### Plasmid pMAD6

Previously described in WO 95/23868, and shown in Figures 1 and 2.

#### Plasmid pMAD1

Two primers, complementary to sequences at the 5' and 3' boundaries of the first intron of the ovine BLG gene, were used to amplify a ~650bp fragment encompassing the entire sequence of intron 1 of the BLG gene from pUCXSRV template. The primers introduce a 5' *Sma*I site and a 3' *Eco*RV site at the ends of the PCR fragment. This fragment was cloned in *Eco* RV digested pBluescriptSK

to which single 3' dATP overhangs were added, using Taq polymerase. This construct was named pSTI1. The orientation of the insert with respect of the multiple cloning site in pSTI1 was determined by restriction digestion.

- 5 The intron 1 sequence was excised from pSTI1 on a 5' *Sma*I -3' *Hind*III fragment, the recessed 3' terminus generated at the *Hind*III end was repaired using Klenow, and the resulting blunt ended fragment was ligated with *Eco*RV digested pMAD to make pMAD1. The correct orientation of the intron fragment with respect to the remainder of the BLG sequences was determined by DNA sequencing. This step  
10 effectively moves the *Eco*RV site to the 3' end of the BLG intron.

#### Plasmid pMAD16

- This was constructed using essentially the same strategy as that described for pMAD1, except that in the final cloning step the BLG intron was ligated with  
15 *Eco*RV cleaved pMAD6 (instead of pMAD) to construct pMAD16.

#### Expression Constructs

##### Plasmid pCORP2 (see WO-A-9211358)

20

- A 1450bp cDNA of the human protein C gene, flanked by *Kpn*I sites, was obtained in the form of plasmid pWAPC2. The cDNA was excised as a *Kpn*I fragment, the 3' overhangs made flush by treatment with T4 DNA polymerase, the fragment gel purified and blunt-end cloned into the *Eco*RV site of pMAD. Orientation was  
25 determined by restriction digest and confirmed by DNA sequencing. This construct is plasmid pCORP2 and contains the human protein C cDNA under the transcriptional control of the  $\beta$ -lactoglobulin gene 5' and 3' flanking sequences. There are no introns.



Plasmid pCORP5

The 1450bp protein C cDNA fragment used in the construction of pCORP2 was placed into pMAD16 to make pCORP5.

5 Plasmid pCORP9

To facilitate the cloning of the protein C cDNA, PC962 (Foster *et al.*, *ibid*), into pMAD, the plasmid was modified to incorporate *EcoRV* sites at the extremities of the protein C cDNA insert. A 769 bp *SstII-PstI* fragment  
10 encompassing the 3' end of PC962 was cloned between the *SstII* and *PstI* sites of pBluescript II SK+ (Stratagene, La Jolla, CA). The fragment was excised with *SstII/EcoRV* and purified. The 5' portion of PC962 was modified by PCR. The sense oligonucleotide primer for this reaction covered the 5' ATG region of the cDNA and provided an *EcoRV* site upstream of this in the product. The  
15 antisense oligonucleotide primer covered the *SstII* site used to generate the *SstII-EcoRV* fragment. The resulting PCR product was digested with *EcoRV* and *SstII* and ligated with the *SstII-EcoRV* 3' fragment and *EcoRV* digested pMAD. The resulting plasmid, designated pCORP9 effectively contained the PC962 cDNA flanked by *EcoRV* sites in an intronless fusion driven by the  $\beta$ -  
20 lactoglobulin promoter.

Plasmid pCORP14

A genomic DNA construct, containing exons I through VIII of the human protein  
25 C gene, was made. This genomic construct, designated GPC10-1, changed the sequence 16 base pairs upstream of the ATG from the native protein C sequence to the  $\beta$ -lactoglobulin sequence and introduced mutations in the propeptide cleavage site located in exon 2, and the two-chain cleavage site located in exon 6, as described below. The construct was assembled using four fragments  
30 designated A, B, C and D and encompassed the protein C gene sequence from

the ATG to a *Bam*HI site in exon VIII, immediately upstream of the stop codon.

The fragments were generated from a human genomic library in Charon 4A phage which was screened with a radiolabeled cDNA probe for human protein C.

The screening of the  $\lambda$  library produced three clones that together mapped the  
5 entire protein C5 gene (Foster *et al.*, 1985, *Proc. Natl. Acad. Sci. USA*, 82: 4673-4677). These clones were designated PC  $\lambda$ 1, PC  $\lambda$ 6 and PC  $\lambda$ 8. Fragment A was a *Not*I to *Eco*RI fragment that contained exons I and II of the genomic sequence and was 1698 bp. A subclone of PC  $\lambda$ 6 contained an *Eco*RI to *Eco*RI fragment and was designated pHCR4.4-1. Using pHCR4.4-1 as a template and  
10 oligonucleotides ZC6303 (5'-ATT TGC GGC CGC CTG CAG CCA TGT GGC AGC TCA CAA GCC TCC TGC-3') and ZC6337 (5'-CAG GAA GGA GTT GGC GCG CTT GCG CCG TTG CAG CAC CTG CTG GGC-3'), a DNA fragment was generated by polymerase chain reaction (PCR). Oligonucleotide ZC6303 changed the sequence 16 based pairs 5' to the ATG sequence from the  
15 native protein C sequence to the equivalent sequence from the  $\beta$ -lactoglobulin gene and introduced a *Not*I site.

Oligonucleotide ZC6337 changed the propeptide cleavage site from Arg-Ile-Arg-Lys-Arg to Gln-Arg-Arg-Lys-Arg. The resulting PCR generated fragment was  
20 digested with *Not*I and *Bss*HII, and a 1402 base pair fragment was gel purified and designated A1. A second fragment was prepared using a  $\lambda$  gtII clone of PC  $\lambda$ 1 as a template with oligonucleotides ZC6306 (5'-CTT CTT CCT GAA TTC TGT TTC TTG C-3') and ZC6338 (5' -CGG ATC CGC AAG CGC GCC AAC TCC TTC C-3') in a polymerase chain reaction. The resulting DNA fragment,  
25 designated A3, was digested with *Bss*HII and *Eco*RI and gel purified, resulting in a 296 base pair fragment.

Fragments A1 and A3 were ligated into the Bluescript II KS+ phagemid vector (Stratagene, La Jolla, CA). The resulting plasmid, designated GPC 2-2, was  
30 digested with *Not*I and *Eco*RI, gel purified and the *Not*I-*Eco*RI DNA fragment

was designated Fragment A.

pCR 2-14 is a subclone which contains an *EcoRI* to *EcoRI* DNA fragment of PC  
λ8 (Foster *et al.*, 1985, *ibid.*). The plasmid was digested with *EcoRI* and *SstI*  
5 and gel purified. The resulting fragment was designated Fragment B.

Plasmid pCR 2-14 was used as a template DNA with oligonucleotides ZC6373  
(5' -AAA GTA AAA AAA GAT CTA AAA ATT TAA C-3') and ZC6305 (5' -  
GTG TCT CGT TTT CTT AAG TGA CTG CGC-3'), which introduced an *AflIII*  
10 site and the RRKR mutation of the native (KR) two-chain cleavage site, in a  
polymerase chain reaction. The resulting PCR-generated fragment was digested  
with *BglIII* and *AflIII* and gel purified, resulting in a 1441 base pair fragment,  
designated E1. Fragment E1 was used in a ligation reaction with  
oligonucleotides ZC6302 (5' -TTA AGA AGA AAA CGA GAC ACA GAA  
15 GAC CAA GAA GAC CAA GTA GAT CCG C-3') and ZC6304 (5' -GGA  
TCT ACT TGG TCT TCT TGG TCT GTG TCT CGT TTT CTT C-3'). These  
oligonucleotides form *AflIII* and *SstII* restriction sites when annealed and were  
ligated to the 3' end of fragment E1, resulting in a fragment with a 5' *BglIII* site  
and a 3' *SstII* site. This fragment was used in a ligation reaction with a *BamHI*-  
20 *SstII* digested Bluescript II KS+ phagemid vector (Stratagene). The resulting  
plasmid was designated GPC 8-5 and digested with *SstI* and *SstII*, generating a  
626 base pair fragment, designated Fragment C.

A fourth fragment was generated by digestion of a genomic subclone (pHCB7-1)  
25 of PC λ8. pHCB7-1 contained a *BglIII* to *BglIII* fragment that encompassed  
exons VI through VIII. pHCB7-1 was digested with *SstII* and *BamHI* and a 2702  
base pair fragment was gel purified. The fragment was designated Fragment D.

A five-part ligation reaction was prepared using *NotI* and *BamHI* digested and  
30 linearized Bluescript II KS+ phagemid vector (Stratagene) with Fragment A (5'

*NotI* to 3' *EcoRI*) that contained exons I and II, Fragment B (5' *EcoRI* to 3' *SstI*) that contained exons III, IV and V, Fragment C (5' *SstI* to 3' *SstII*) that contained the 5' portion of exon VI and Fragment D (5' *SstII* to *BamHI*) that contained the remaining 3' portion of exon VI and exons VII and VIII.

5

The resulting DNA was 8950 base pairs and designated GPC 10-1.

GPC10-1 was originally generated with BLG sequences and a *NotI* site upstream of the ATG initiator codon and modifications to both cleavage sites. A clone, designated pPC12/BS, was generated to ensure that the 5' *NotI* site of GPC10-1 would not introduce secondary structure into mRNA molecules that could hinder translation. pPC12/BS was generated using PCR amplification of a 1 kb *NotI*-*ScaI* fragment that covered the 5' region of the protein C gene and contained the wild-type ATG codon environment. This introduced an *EcoRV* site immediately downstream of the *NotI* site, adjacent to the ATG codon, and a *BamHI* site was incorporated 3' of the *ScaI* site to facilitate cloning. Following a *NotI/BamHI* digestion, the PCR product was cloned into *NotI/BamHI* digested Bluescript II KS+ phagemid vector (Stratagene). The *NotI-EcoRV-ScaI* fragment present in pPC12/BS was excised, purified and ligated to GPC10-1, which had been linearized with *NotI* and partially digested with *ScaI* (the pUC ampicillin gene has an internal *ScaI* site). The resulting clone was designated GPC10-2 and possesses an *EcoRV* site immediately upstream of the ATG initiator codon. GPC10-1 and GPC10-2 both terminated at the final *BamHI* site in exon VIII of the protein C gene. To reconstitute the 56 bp of sequence, ending at the termination codon, two oligonucleotides were synthesized with flanking *BamHI* (5') and *BglII* (3') restriction sites. Following annealing of the oligonucleotides, the product was cloned into *BamHI* digested pBST+ to generate plasmid pPC3'. pBST+ is a derivative of pBS (Stratagene) with a new polylinker. The addition of the polylinker added *BglII*, *XhoI*, *NarI* and *ClaI* restriction sites from the vector polylinker downstream of the destroyed *BglII* site of the oligonucleotide

30

construct.

The *NotI*-*Bam*HI fragment of GPC10-1 was subcloned into *NotI*/*Bam*HI digested pPC3' to add 3' coding sequences of protein C, the TAG termination codon followed by *Bgl*III-*Xho*I-*Nar*I-*Cla*I. The 3' region of the protein C gene beginning with the *Eco*RV site in intron V was excised from this plasmid on an *Eco*RV-*Cla*I fragment.

The *Eco*RV-*Eco*RV fragment from GPC10-2, covering the 5' portion of the protein C gene, and the 5' above *Eco*RI-*Cla*I fragment covering the 3' portion of the protein C gene were combined between the *Eco*RV and *Cla*I sites of pMAD6 to generate pCORP13. This effectively placed a genomic portion of the protein C gene with modified propeptide and two-chain cleavage site under the control of the  $\beta$ -lactoglobulin promoter.

A further genomic construct was generated from pCORP13 which contained only the modified two-chain cleavage site. This was achieved using PCR amplification to modify two fragments which result in restoration of the coding capability of exon 2 from the mutant Gln-Arg-Arg-Lys-Arg to the wild-type Arg-Ile-Arg-Lys-Arg. pCORP13 was used as template for these reaction. The first fragment was 1.3kb, which encompassed the 5' end of the protein C gene up to the *Bam*HI site in exon 2. For this reason, the sense primer was designed to add a *Hind*III site 5' to the *Eco*RV site proximal to the ATG initiation codon. The antisense primer was designed to restore the wild-type sequences in exon 2, which included a restored *Bam*HI site. A second fragment of 0.2kb from the *Bam*HI site in exon 2 to the *Xho*I site in intron 2, was amplified. The two fragments were combined in pGEMII (Promega, Madison, WI) to generate pGEMOC1.5. A 7.5kb *Xho*I fragment from pCORP13 was ligated to *Xho*I digested pGEMPC1.5 to generate a complete protein C genomic sequence covering exons 1-8 with a wild-type propeptide cleavage site and a modified two-chain cleavage site. The plasmid

was designated pGEMPC14. The sequence was excised from pGEMPC14 as a *HindIII/SaII* fragment. The DNA termini was repaired using a Klenow reaction and the fragment was blunt-end ligated into *EcoRV* digested pMAD6 to generate pCORP14.

5

#### Plasmid pCORP16

The modified protein C cDNA (PC962) was excised from the plasmid pCORP9 (*see above*) as an *EcoRV* fragment and ligated with *EcoRV* pMAD6. The resulting construct has been named pCORP16.

10

#### The Vector pCASMAD6

#### Plasmid pE'10

15 The Bovine  $\beta$ -Casein intron 1 (BBCI 1; BOVCAS1 (5'-AGG CCT ATT CAG CTC CTC CTT CAC TTC TT-3') and BOVCAS2 (5'-GAT ATC GGC TCT CAA TTC CTG GGA ATG GG-3') approximately 2 Kbp) was PCR amplified from dairy cow DNA. The 5' primer incorporates a *StuI* site and the 3' primer incorporates an *EcoRV* site. The purified 2 Kb fragment was cloned into the  
20 pGEM-T vector (Promega) to give construct pE'10.

#### Plasmid pCASMAD6

pMAD6 was modified by inserting a linker, containing *Spe I/Not I/Sac II* sites, into the *EcoRV* site. Both orientations of the linker were obtained and thus two new  
25 cloning vectors were obtained. These were called pMAD6/STOPS (5' *SacII/NotI/SpeI* 3') and pMAD6/SPOTS (5' *SpeI/NotI/SacII* 3').

BBCI 1 was excised from pE'10 on a *SacII* and *SpeI* partial (due to an internal *SpeI* site in the  $\beta$ -Casein intron) and cloned into *SacII/SpeI* digested pMAD6/SPOTS.

30 The new vector was called pCASMAD6.

Plasmid pCORI69

The modified protein C cDNA (PC962) was excised from the plasmid pCORP9 (see above) as an *EcoRV* fragment and ligated with *EcoRV* digested pCASMAD6.

This places the AUG translation start downstream with respect to the  $\beta$ -casein  
5 intron sequence. The resulting construct was named pCOR169.

The Vector pACTMAD6Plasmid pGEM-AI

10 Two primers, (Sequences ACTP1 5'-AGG CCT AGT GCC TGC CAC CAG  
CGC CAG CC-3' ACTP2 5'-GAT ATC CCT GGC ACA GCT TTG TGT GGT  
TC-3') complementary to the opposing strands of the 3' end of the first exon and  
the 5' end of the second exon of the murine cardiac actin gene respectively, were  
used in a PCR reaction to amplify a 0.8 Kb fragment encompassing the intervening  
15 sequences from a template of mouse genomic DNA. The two primers introduced a  
5' *SnaBI* and a 3' *EcoRV* restriction site at the ends of the PCR product. This  
DNA fragment was cloned in pGEM-T to give a construct which was named  
pGEM-AI. DNA sequence analysis confirmed that the sequence of the amplified  
product beyond the primers matched that published for the murine beta actin gene.

20

Plasmid pACTGMAD6

The actin intron 1 sequence was excised from pGEM-AI on a 5' *SnaBI*- 3'*EcoRV*  
fragment which was then ligated with *EcoRV* digested pMAD6 to give vector  
pACTMAD6. This cloning step effectively moves the *EcoRV* site from the 3' end  
25 of the BLG promoter downstream to the 3' end of the actin gene intron segment.

Plasmid pCORI70

The modified protein C cDNA (PC962) was excised from the plasmid pCORP9 as  
an *EcoRV* fragment and ligated with *EcoRV* digested pACTMAD6. This places  
30 the AUG translation start downstream with respect to the actin intron sequence.

i

The resulting construct was named pCOR170.

### The Vector pBOB

#### 5 Plasmid pBOB

PCR primers were designed to amplify the region of the ovine  $\beta$ -casein gene from exon 1 to exon 2 (BOB1: 5'-CGG GAT CCG TCG ACC ATT CAG CTT CTC CTT CAC TTC TTC TC-3'; BOB2: 5'-CGG GAT CCG GGT CCC TAC GTA GGC TCT CGA TTC CTG TGA ATG GGA-3'). The size of this product is  
10 2.1Kbp and has, engineered into it, the sites *Bam*HI/*Sal*I at the 5' end and *Bam*HI/*Ppu*MI/*Sna*BI at the 3' end.

The construction of pBOB took place in three steps: the 2.1Kbp PCR fragment (above) was blunt-end cloned into the *Eco*RV site of pBSIISK+ (Stratagene) to  
15 give the plasmid p $\beta$ -casEx1/2. The 6.5Kbp *Ppu*MI fragment from the ovine  $\beta$ -casein gene, containing exons 7 to 9 and 3' flanking sequences, was cloned into the now unique *Ppu*MI site of pO $\beta$ -caseEx1/2. A clone was obtained with the 6.5Kbp fragment in its natural orientation with respect to the first intron and this clone was named pBOB $\Delta$ prom.

20

Previously, a *Xho*I linker had been cloned into the *Eco*RV site of pMAD6 and the modified plasmid named pMADX. Finally, the ovine BLG promoter from pMAD6X was cloned into the *Sal*I site of pBOB $\Delta$ prom as a *Sal*I/*Xho*I fragment giving rise to pBOB.

25

#### Plasmid pCORB71

The modified protein C cDNA (PC962) was excised from the plasmid pCORP9 as an *Eco*RV fragment and ligated with *Eco*RV digested pBOB. This places the AUG translation start downstream with respect to the actin intron sequence. The  
30 resulting construct was named pCORB71.



Example 1

Analysis of constructs of the present invention in the expression of protein C in transgenic animals. Results are shown in Tables 1 and 2.

5    Generation of transgenic animals

Transgenic mice were generated as described in Prunkard *et al* (Nature Biotechnology, 14:867-871, 1996).

Protein C Assays

- 10    Human protein C in the milk of transgenic animals was assayed according to the following procedure:

**Protein C Standard**

- 15    Purified human Protein C stored at 50µg/ml in Phosphate Buffered Saline (PBS)/1% bovine Serum Albumin (BSA) at 20 °C. Dilute to 500ng/ml in blocking buffer for use. Standard curve range of ELISA is 3.9-125ng/ml.

**Blocking buffer**

- 1X PBS  
20    5% Milk powder  
0.01% Tween 20

**Wash buffer**

- 1X PBS  
25    0.05% Tween 20

**Coating Antibody**

Dako Rabbit Anti-human Protein C antibody diluted to 10µg/ml in PBS

**Detection Antibody**

Dako Rabbit Anti-human Protein C antibody Peroxidase conjugate.

Dilute 1/5000 in blocking buffer.

**5 Substrate**

TMB 1 Component Peroxidase substrate

**Stop Solution**

0.2M Sulphuric acid

10

**Method**

1. Coat Costar High Binding capacity 96 well plate with 150 $\mu$ l/well of coating antibody. Incubate in a damp box o/n in fridge.
2. Wash wells with wash buffer (3X 200 $\mu$ l/well).
- 15 3. Load 100 $\mu$ l blocking buffer into wells.
4. Dilute samples appropriately in blocking buffer.  
Reference human plasma is used as a protective control at 1/40 dilution.
5. Load standard and samples into plate by columns with doubling dilution (100 $\mu$ l per well). Standard is loaded in rows 1 and 12 in duplicate.
- 20 6. Incubate for 2 hours in damp box in fridge.
7. Wash plate with wash buffer (3X 200 $\mu$ l/well).
8. Load 100 $\mu$ l/well peroxidase conjugate. Incubate in damp box for 2 hours in fridge.
9. Wash plate with wash buffer (3X 200 $\mu$ l/well). Drain plate.
- 25 10. Add 100 $\mu$ l/well of substrate and leave for 5 minutes.  
Stop reaction by addition of 100 $\mu$ l/well of stop solution.
11. Read plate on plate reader with 650nm filter.
12. Plot standard curve using mean of duplicates (O.D v.log concentration PC) and calculate regression line equation. Use equation to calculate sample values.
- 30 values. Data handling can be performed by PC assay programme on Dynex

Revelation software.

### Example 2

- 5 Expression of AAT from cDNA in constructs according to the present invention.

Transgenic mice were prepared as in Example 1. Analysis of AAT in the milk of transgenic mice was according to standard procedures, for example as described in Wright, G., Carver, A., Cottom, D., Reeves, D., Scott, A., Simons, J.P.,  
10 Wilmut, I., Garner, I., and Colman A., 1991. High level expression of active human  $\alpha 1$  antitrypsin in the milk of transgenic sheep. Bio/Technology 9: 830-834.

Results are shown in Table 3.

15 Example 3.

Expression of antibody fragment from constructs according to the present invention.

- 20 Constructs pMAD6 and pCASMAD6 were prepared incorporating DNA encoding an antibody binding fragment to give constructs pMAD6-AB and pCASMAD6-AB. The constructs were used to obtain transgenic mice according to Example 1. Expression of the antibody fragment was determined by standard protocols.

- 25 No expression of the antibody fragment was found in the transgenic mice with pMAD6-AB. Levels ranging from 0 to 129  $\mu\text{g/ml}$  were found in pCASMAD6-AB mice.

The results are given in Table 4.

Example 4

Expression of IgG from constructs according to the present invention.

Constructs pMAD6 and pCASMAD6 were prepared incorporating DNA encoding  
5 IgG to give constructs pMAD6-IgG and pCASMAD6-IgG. The constructs were  
used to obtain transgenic mice according to Example 1. Expression of IgG in the  
mice milk was determined by standard ELISA protocol.

Results are given in Table 5.

10

Example 5

Expression of adhesion molecule (soluble) from constructs according to the present  
invention.

15

Constructs pMAD6 and pCASMAD6 were prepared incorporating DNA encoding  
a soluble adhesion molecule (SAM) to give constructs pMAD6-SAM and  
pCASMAD6-SAM. Transgenic mice were prepared according to Example 1.

20 The expression level range ( $\mu\text{g/ml}$ ) in pCASMAD6-SAM transgenic mice was up  
to 500. The maximum level detected in the pMAD6-SAM transgenic mice was 80.

Example 6

25 Expression of collagen cDNA from constructs according to the present invention.

The CASMAD6 vector was used. Collagen cDNA (human truncated pro-collagen  
 $\alpha 2(1)$  homotrimer) was inserted as the DNA for the protein of interest. This was  
coinjected with two transgenes expressing  $\alpha$  and  $\beta$  subunits of prolyl 4-  
30 hydroxylase, an enzyme for the post-translational modification of procollagen.

Transgenic animals were obtained as in Example 1. Determination of collagen expression in mouse milk was as described according to standard protocols and described in WO97/08311.

5

11 transgenic lines bearing the collagen cDNA construct and the two prolyl 4-hydroxylase transgenes were analysed. Three lines were found to express procollagen  $\alpha 2(1)$  homotrimer protein. The amount of collagen present was estimated by measurement of hydroxyproline content and Western analysis in  
10 comparison with bovine collagen standard. Three independently derived mouse lines were found to express detectable amounts of collagen. The levels present in milk of these three lines was estimated as:  $10\mu\text{g/ml}$ ,  $30\mu\text{g/ml}$ , and  $120\text{--}240\mu\text{g/ml}$ . Collagen protein was absent from the milk of non-transgenic mice. Milk from the highest expressing line was analysed further and the procollagen present was found  
15 to have formed a correctly aligned triple helical molecule.

These results demonstrate secretion of relatively high levels of recombinant procollagen in the milk of transgenic mice by expression of cDNA under the control of the  $\beta$ -lactoglobulin promoter.

20

From the foregoing it will be appreciated that although specific embodiments of the invention have been described herein for purposes of illustration, various modifications may be made without deviating from the spirit and scope of the invention. Accordingly, the invention is not limited. All documents and papers  
25 cited or mentioned herein are fully incorporated by reference.

**TABLE 1**

**Data supporting the improved expression from constructs according to the invention**

5

Logic		CDNA	Construct
Addition of an intron does NOT result in high expression	PCORP3	Wild type*	pMAD6
	PCORP16	Mutant*	pMAD6
Addition of ANY intron into the 5'UTR is insufficient	PCORP4	wild type	pMAD1
	PCORP5	wild type	pMAD16
Addition of the naturally occurring $\beta$ -casein 5'UTR intron results in high expression of ANY cDNA	PCORI69	mutant	pCASMAD6
Addition of ANY naturally occurring 5'UTR intron results in high expression of ANY cDNA	PCORI70	mutant	pACTMAD6

\*refers to different protein C cDNAs

Table 2: Protein C Expression data

Plasmid	Type	Intron	3' Intron	Expression Range (ug/ml)	Notes
pCORP2	Wild type cDNA	none	none	0.0	
pCORP2 (rescued)	Wild type cDNA	none	none	0.0 - 80.0	Co-injected with o-BLG
pCORP3	Wild type cDNA	none	o-BLG 16	1.1 - 27.4	
pCORP3 (rescued)	Wild type cDNA	none	o-BLG 16	5.9 - 55.8	Co-injected with o-BLG
pCORP5	Wild type cDNA	o-BLG I1	o-BLG 16	0.0	
pCORP16	Mutant cDNA	none	o-BLG 16	0.0 - 172.0	Expressors = 10/20
pCORP9	Mutant cDNA	none	none	0.0	
pCORP9 (rescued)	Mutant cDNA	none	none	0.0 - 2.5	Co-injected with o-BLG
pCORP14	Mutant genomic	none	o-BLG 16	1.7 - 192.0	Expressors = 8/8
pCORI69	Mutant cDNA	Bov $\beta$ -casein I1	o-BLG 16	0.0 - 186.9	Expressors = 6/7
pCORI70	Mutant cDNA	Mus $\beta$ -casein I1	o-BLG 16	0.0 - 228.0	Expressors = 6/8
pCORB71	Mutant cDNA	Ov $\beta$ -casein I1	o- $\beta$ -casein	0.0 - 133	Expressors = 19/20

**TABLE 3****Expression of cDNA Constructs**

Vector	Gene	BLG rescue	Founders expres. > 10 µg/ml	Protein µg/ml mean
PCASMAD6	AAT	no	7/21	1320
PMAD6	AAT	no	1/20	10



**TABLE 4**

5

**pMAD6-AB**

Mouse No.	EXPRESSION ( $\mu\text{g/ml}$ )
#33	0
#34	0
#35	0
#57	0

10

**pCASMAD6-AB**

Mouse No.	EXPRESSION ( $\mu\text{g/ml}$ )
#14	11.8 + 0.75
#55	36.17 + 0.175
#58	21.76 + 1.46
#60	20.8 + 0.67
#69	0
#72	1.9 + 0.01
#75	2.5 + 0.01
#81	0
#102	0.44 + 0.03
#110	0
#194	54.31 + 7.1
#195	129.41 + 1.38
#205	8.38 + 2.41
#216	45.6 + 0.12

**TABLE 5**

## Expression Levels of IgG in Transgenic Mouse Milk

5

TG exp. vector Generation /-IgG		IgG Expression level -ELISA- (mg/l)
CASMAD6	G0	< 0.05
CASMAD6	G0	80 - 102.4
CASMAD6	G0	22.4 - 35.8
CASMAD6	G0	< 0.05
CASMAD6	G0	< 0.05
MAD6	G0	0.5 - 0.8
MAD6	G0	2.3 - 3.5
MAD6	G0	< 0.05
MAD6	G0	< 0.05
MAD6	G0	< 0.05
MAD6	G0	< 0.05
MAD6	G0	< 0.05
MAD6	G0	< 0.05
CASMAD6	G1	0.18 - 0.2
CASMAD6	G1	0.5 - 0.6
CASMAD6	G1	15 - 16

**CLAIMS**

1. A nucleic acid expression construct comprising:

- (a) a promoter;
- 5 (b) an intron whose natural position is within the 5'-untranslated region of a gene from which it is derived;
- (c) a coding sequence; and
- (d) a 3'-flanking sequence

wherein the intron (b) is not derived from the same gene as that from which either  
10 the promoter (a) or the protein-coding sequence (c) is derived.

2. An expression construct as claimed in claim 1 wherein the promoter is a gene promoter which drives expression of the coding sequence (c) in mammalian cells, in particular, a milk protein promoter.

15

3. An expression construct as claimed in claim 1 or claim 2 wherein the intron (b) is the first intron from a gene where the intron is naturally located entirely within the 5' untranslated region of the gene.

20 4. An expression construct as claimed in 1, 2, or 3, wherein the intron (b) is the first intron from a gene which is a member of a family of genes where the intron is naturally located entirely within the 5' untranslated region of the gene.

5. An expression construct as claimed in any one of claims 1 to 4 wherein the  
25 intron is from the casein gene family.

6. An expression construct as claimed in any one of claims 1 to 4 wherein the intron is from the actin gene family.

30

7. An expression construct as claimed in any one of claims 1 to 6 wherein the 3' flanking sequence is any sequence which supports the correct transcription termination, mRNA 3' end processing, mRNA stabilisation, mRNA transport from the nucleus to cytoplasm and mRNA translation.

5

8. An expression construct as claimed in any one of claims 1 to 7 wherein the 3' flanking sequence is a poly-A site or a  $\beta$ -lactoglobulin gene 3'-sequence beginning 3' to the natural  $\beta$ -lactoglobulin stop codon and continuing to at least about 50 bases 3' of the poly-A site.

10

9. An expression construct as claimed in any one of claims 1 to 7 wherein the 3'-flanking sequence is a  $\beta$ -casein gene 3'-sequence beginning 3' to the natural  $\beta$ -casein stop codon and continuing to at least about 50 bases 3' of the poly-A site.

15

10. A process for the preparation of a host organism the process comprising introducing an expression construct, as claimed in any one of claims 1 to 9, into a suitable organism.

20

11. A process as claimed in claim 10 wherein the suitable organism is a prokaryote, a fungi, a plant, an animal, or a eukaryotic cell.

12. A process as claimed in claim 11 wherein the animal is a non-human mammal.

25

13. A host organism incorporating a DNA expression construct as claimed in any one of claims 1 to 9.

14. A host organism as claimed in claim 13 which is a prokaryote (eg. *E.coli*), a fungi, a plant, an animal or a eukaryotic cell.

30

15. A host organism, as claimed in claim 14, wherein the animal is a non-human mammal.

16. A process of preparing a protein, the process comprising allowing an  
5 expression host to express a DNA expression construct as claimed in any one of claims 1 to 9.

17. A process as claimed in claim 16, further including a process of purifying the protein.

10

18. A process as claimed in claim 15 or claim 16 wherein the protein is protein C, fibrinogen, AAT or collagen.

19. A process as claimed in claim 16 or claim 17 wherein the expression host is  
15 a prokaryote, a fungi, a plant, an animal or a eukaryotic cell.

20. A process, as claimed in claimed 19, wherein the animal is a non-human mammal.

20 21. A protein prepared by a process as claimed in any one of claims 16 to 20.

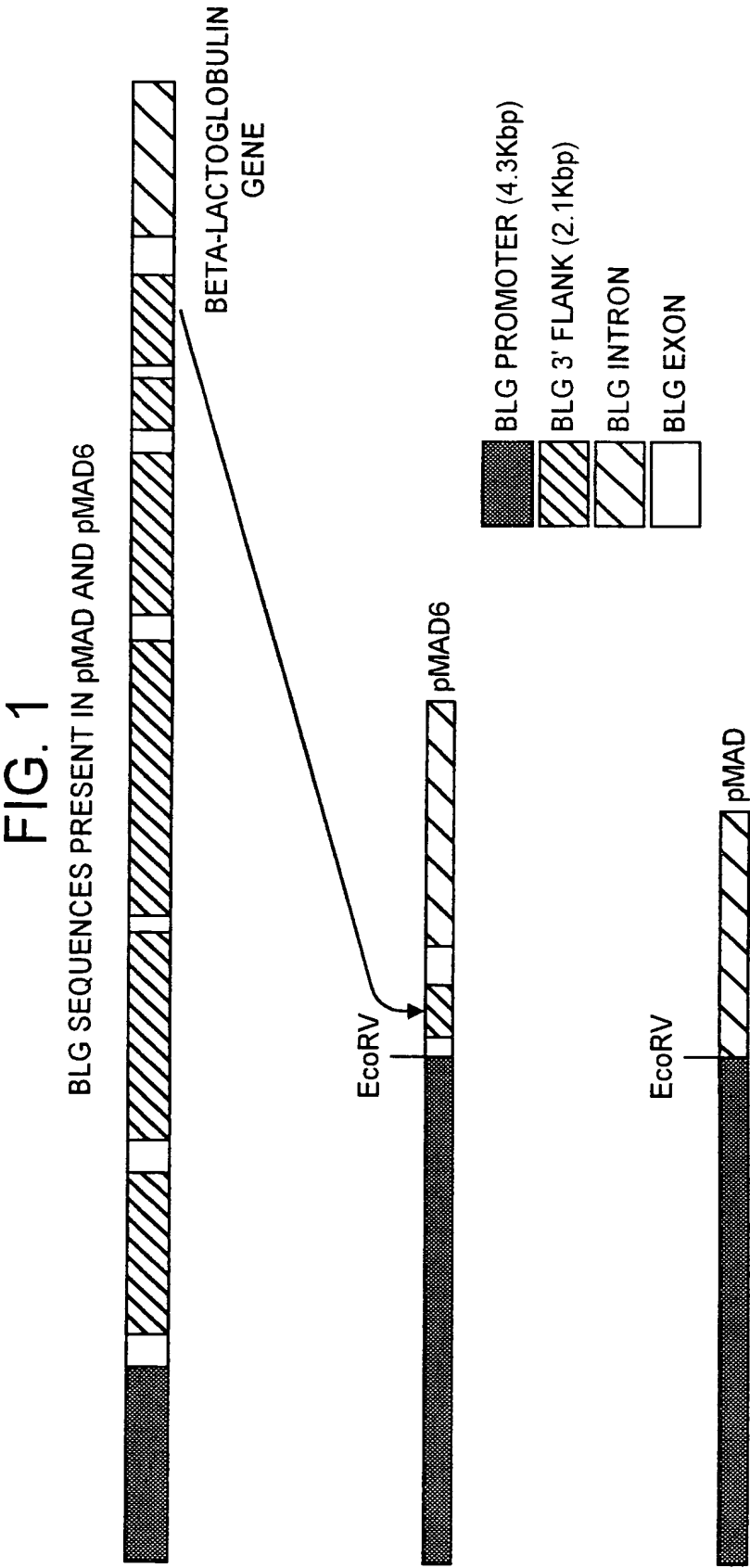
22. The use of a nucleic acid expression construct comprising a promoter, an intron whose natural position is within the 5' untranslated region of a gene from which it is derived, a coding sequence and a 3' flanking sequence to obtain a  
25 transgenic host.

23. The use of nucleic acid construct comprising a promoter, an intron whose natural position is within the 5' untranslated region of a gene from which it is derived, a coding sequence and a 3' flanking sequence to increase the likelihood  
30 of expression of the coding sequence from a transgenic host which incorporates

the nucleic acid construct.

24. A process for improving the number of transgenic hosts which express a transgene coding sequence, the process comprising introducing into the host a nucleic acid construct comprising a promoter, an intron whose natural position is  
5 within the 5' untranslated region of a gene from which it is derived, a coding sequence and a 3' flanking sequence.

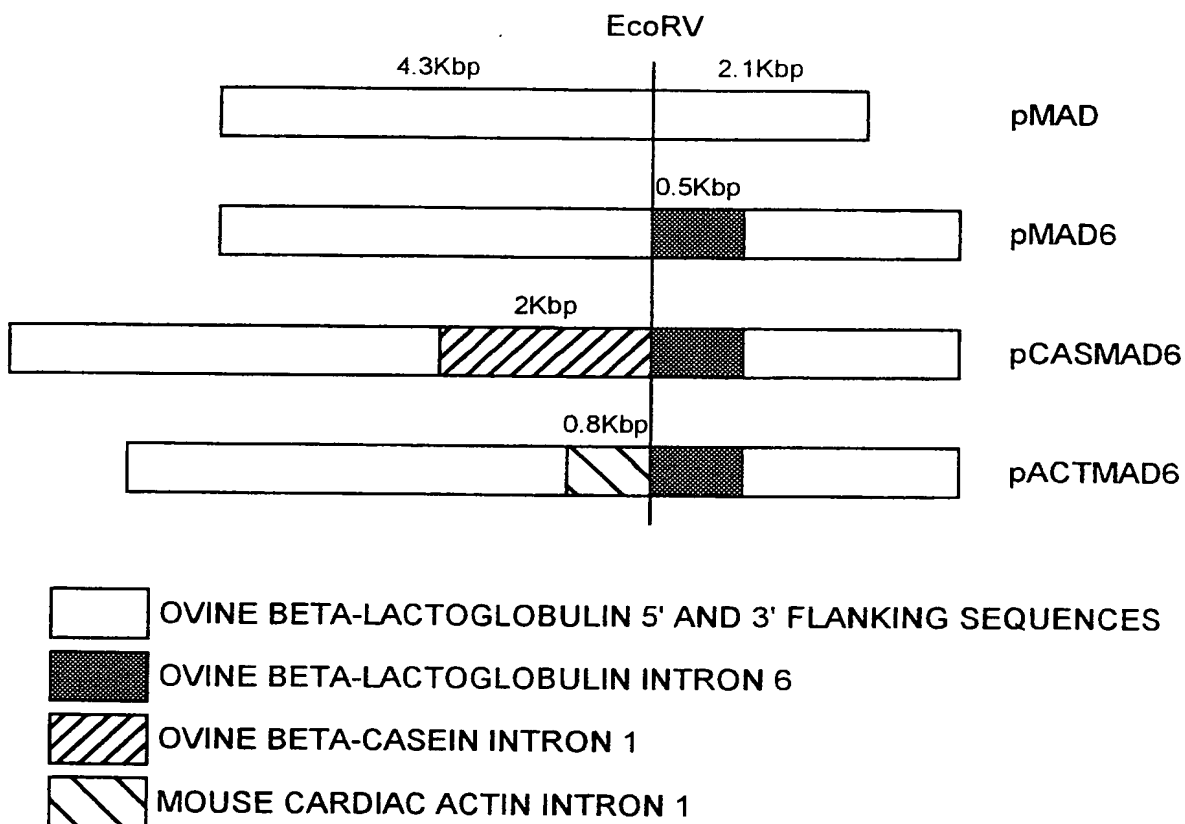
25. The use as claimed in any one of claims 22 or 23 or a process as claimed in claim 24 wherein the construct is as set out in any one of claims 1 to 9.



2 / 7

**FIG. 2**

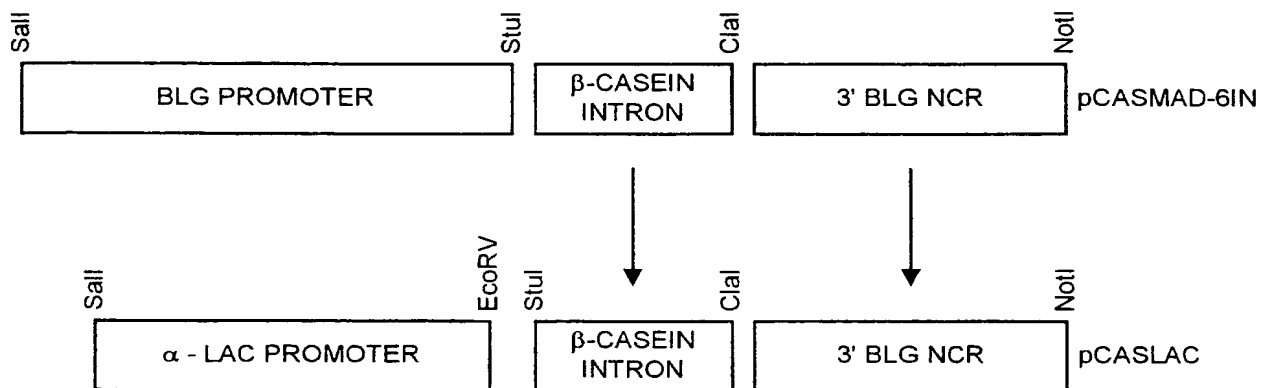
ORIGIN OF SEQUENCES IN VARIOUS VECTORS EMPLOYED





3 / 7

## CLONING OF CASLAC



## TRANSGENE CONSTRUCT

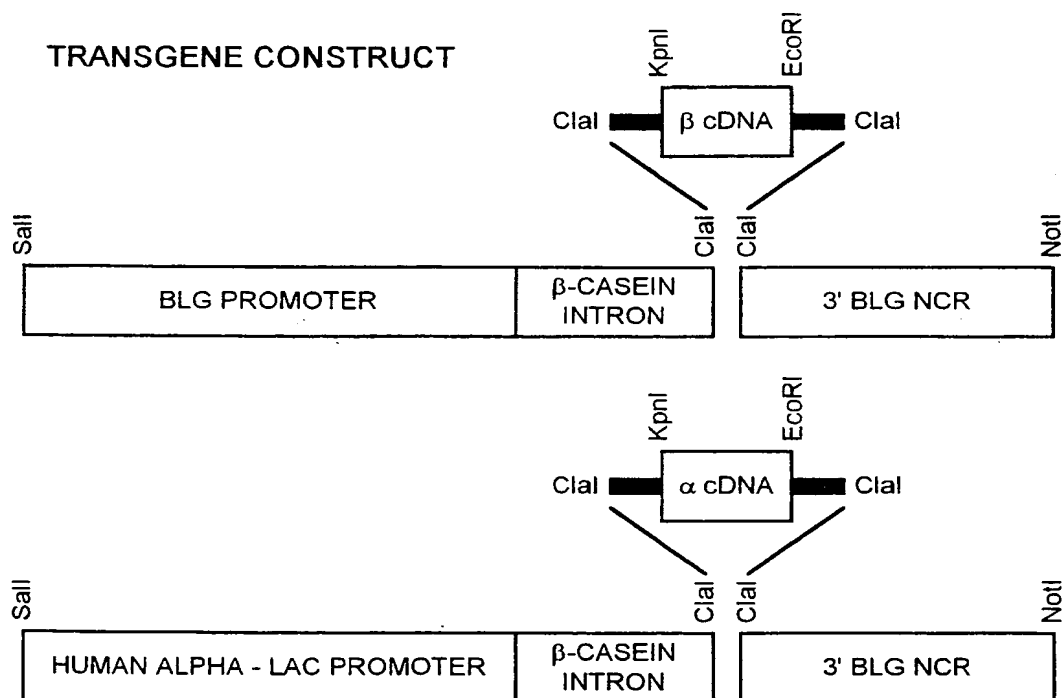
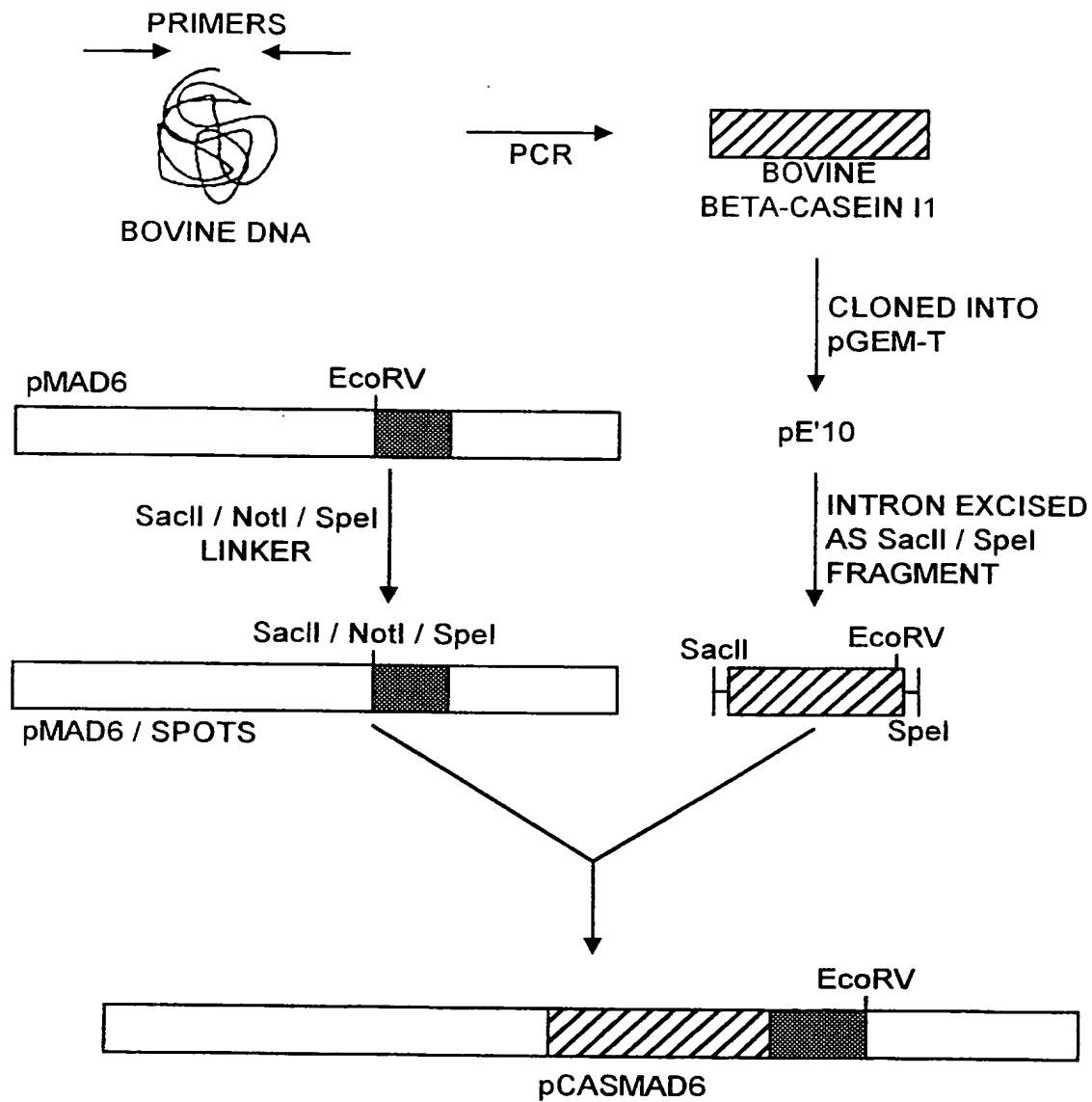


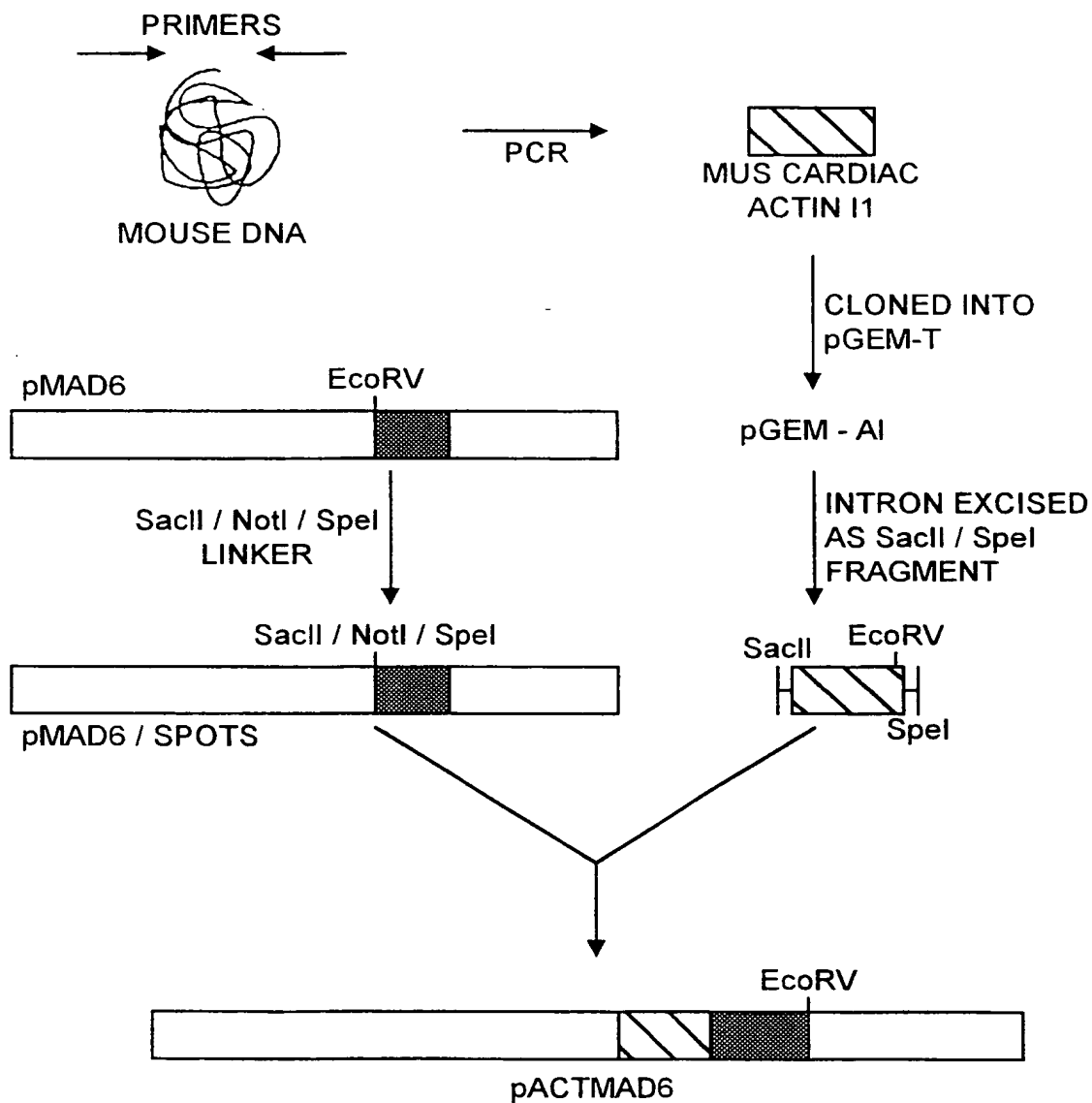
FIG. 3

4 / 7



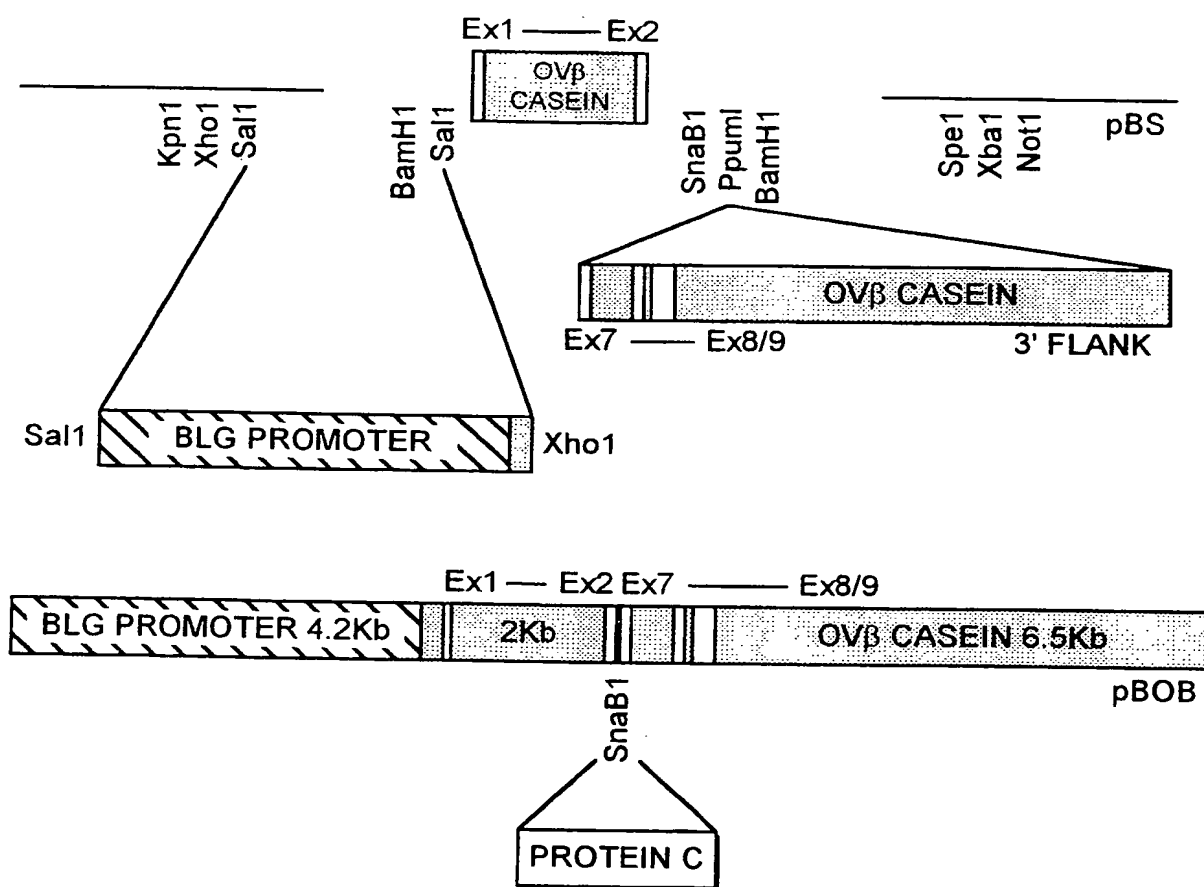
**FIG. 4**  
CONSTRUCTION OF pCASMAD6

5 / 7



**FIG. 5**  
CONSTRUCTION OF pACTMAD6

6 / 7



**FIG. 6**  
CONSTRUCTION OF pBOB

7 / 7

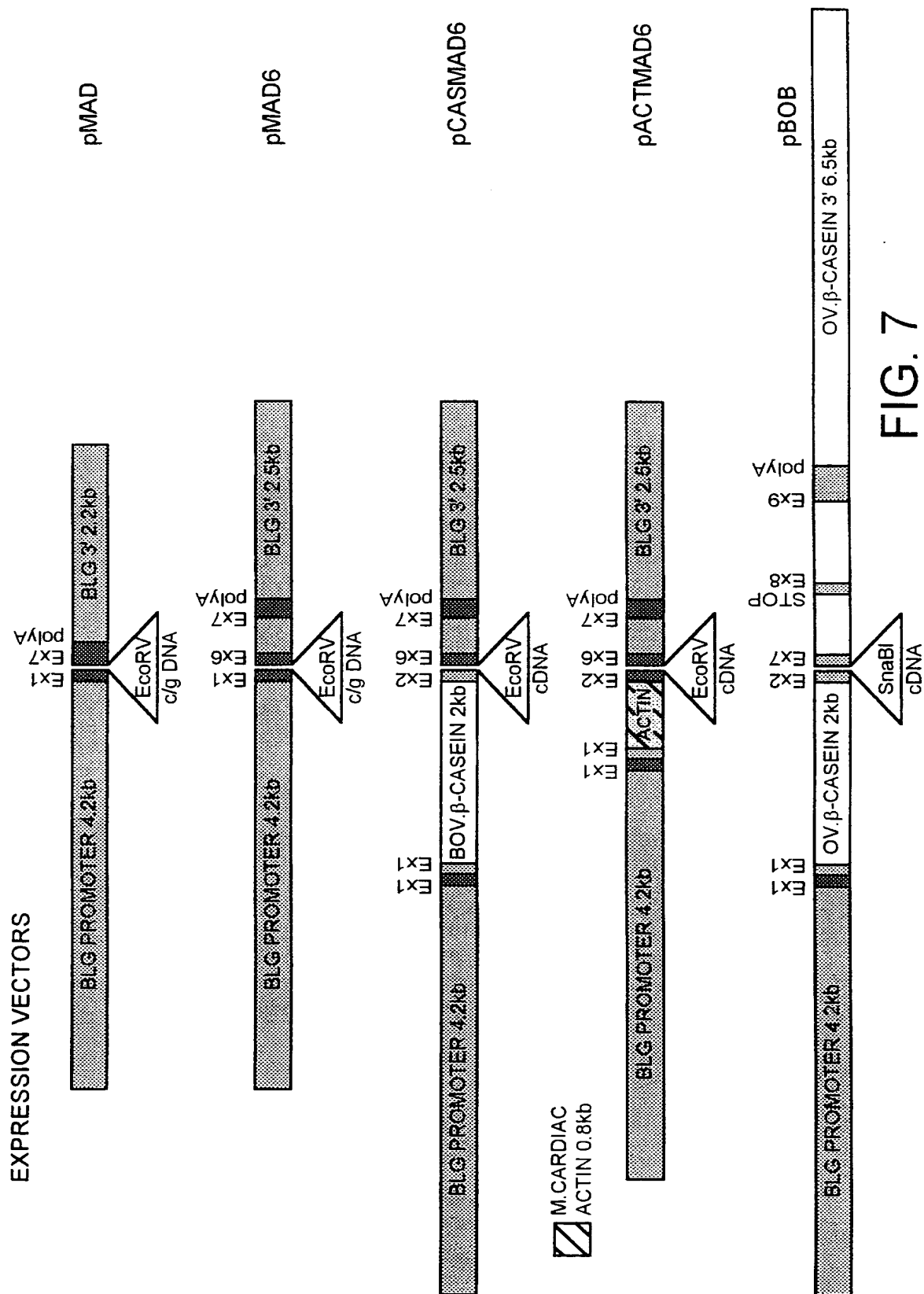


FIG. 7

# INTERNATIONAL SEARCH REPORT

Inte: onal Application No

PCT/GB 98/02130

## A. CLASSIFICATION OF SUBJECT MATTER

IPC 6 C12N15/00 C12N15/12 C12N15/85 C12N5/10 C07K14/705  
C07K14/78 C07K14/81 C07K16/00 A01K67/027

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC 6 C12N C07K A01K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 279 582 A (BAYLOR COLLEGE MEDICINE) 24 August 1988 see page 5, column 7, line 30 - line 34; figures 1A,1B,,8 ----	1-25
X	EP 0 390 592 A (KYOWA HAKKO KOGYO KK ;CENTRAL INST FOR EX ANIMALS (JP)) 3 October 1990 see page 2, line 46 - line 49 see page 3, line 57 - page 4, line 27 see page 4, line 28 - line 55; claims 1-21; figures 1-7 ----	1-5,7, 10-17, 19-25
X	WO 93 25567 A (GENPHARM INT ;DEBOER HERMAN A (NL); STRIJKER REIN (NL); HEYNEKER H) 23 December 1993 see claims 63-83; figures 5A-5C,7E ----	1-5,7, 10-17, 19-25
	-/-	

☒ Further documents are listed in the continuation of box C.

☒ Patent family members are listed in annex.

### \* Special categories of cited documents :

- "A" document defining the general state of the art which is not considered to be of particular relevance
- "E" earlier document but published on or after the international filing date
- "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
- "O" document referring to an oral disclosure, use, exhibition or other means
- "P" document published prior to the international filing date but later than the priority date claimed

- "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
- "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
- "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.
- "&" document member of the same patent family

Date of the actual completion of the international search

16 December 1998

Date of mailing of the international search report

22/12/1998

Name and mailing address of the ISA

European Patent Office, P.B. 5818 Patentlaan 2  
NL - 2280 HV Rijswijk  
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,  
Fax: (+31-70) 340-3016

Authorized officer

Hornig, H

# INTERNATIONAL SEARCH REPORT

International Application No

PCT/GB 98/02130

## C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	EP 0 771 874 A (ISRAEL STATE) 7 May 1997  see the whole document ---	1,3,4,7, 10-16
X	WO 96 29411 A (HEARTLEIN MICHAEL W ;TRANSKARYOTIC THERAPIES INC (US); HAUGE BRIAN) 26 September 1996 see figures 1-3 ---	1,3,4,7, 10-16
X	WO 93 09236 A (BAYLOR COLLEGE MEDICINE) 13 May 1993  see page 14, line 21 - line 32 ---	1,3,4,6, 7,10-17, 19-25
X	WO 93 22430 A (BAYLOR COLLEGE MEDICINE ;UNITED STATES OF AMERICA (US)) 11 November 1993 see page 10, line 10 - line 23; claims 1-54; figures 1,4-8,10; example 1 ---	1,3,4,7, 10-25
X	EP 0 174 608 A (UNIV LELAND STANFORD JUNIOR) 19 March 1986  see page 3, line 23 - page 5, line 2; claims 1-15; figures 1,2 ---	1,3,4,6, 7,10-17, 19-25
X	WO 89 01036 A (CELLTECH LTD) 9 February 1989  see page 3, line 6 - page 5, line 14; figure 1 ---	1,3,4,7, 10-17, 19-25
A	WO 97 20043 A (ZYMOGENETICS INC ;PPL THERAPEUTICS (GB)) 5 June 1997 see page 18, line 33 - page 21, line 35 ---	1-25
A	WO 95 23868 A (ZYMOGENETICS INC ;PHARMACEUTICAL PROTEINS LTD (GB)) 8 September 1995 cited in the application see the whole document ---	1-25
A	WO 88 00239 A (PHARMACEUTICAL PROTEINS LTD) 14 January 1988 cited in the application see the whole document ---	1-25
P,X	EP 0 832 981 A (PHARMING BV) 1 April 1998 see page 5, column 7, line 30 - line 34; claims 1-13; figures 1A,1B,8 -----	1-25

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/GB 98/02130

## Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☒ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:  
Remark: Although claims 23 and 25 (as far as in vivo methods are concerned) is(are) directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.
2. ☐ Claims Nos.:  
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

### Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
- ☐ No protest accompanied the payment of additional search fees.



# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/GB 98/02130

Patent document cited in search report		Publication date	Patent family member(s)	Publication date
EP 0279582	A	24-08-1988	AU 618524 B AU 1178488 A AU 661696 B AU 1941092 A CA 1321764 A EP 0832981 A JP 63309192 A US 5565362 A US 5304489 A	02-01-1992 18-08-1988 03-08-1995 24-09-1992 31-08-1993 01-04-1998 16-12-1988 15-10-1996 19-04-1994
EP 0390592	A	03-10-1990	JP 2261386 A CA 2013453 A	24-10-1990 30-09-1990
WO 9325567	A	23-12-1993	AU 4537393 A EP 0652889 A JP 8504562 T US 5633076 A US 5741957 A	04-01-1994 17-05-1995 21-05-1996 27-05-1997 21-04-1998
EP 0771874	A	07-05-1997	AT 171724 T DE 69600719 D	15-10-1998 05-11-1998
WO 9629411	A	26-09-1996	US 5733746 A AU 5362596 A CA 2215618 A EP 0815232 A	31-03-1998 08-10-1996 26-09-1996 07-01-1998
WO 9309236	A	13-05-1993	US 5298422 A AU 660751 B AU 3124693 A CA 2122617 A EP 0635060 A JP 7500967 T PT 101042 A US 5756264 A	29-03-1994 06-07-1995 07-06-1993 13-05-1993 25-01-1995 02-02-1995 28-02-1994 26-05-1998
WO 9322430	A	11-11-1993	AU 4220793 A CA 2134675 A EP 0652948 A JP 8503844 T	29-11-1993 11-11-1993 17-05-1995 30-04-1996
EP 0174608	A	19-03-1986	JP 61108389 A	27-05-1986
WO 8901036	A	09-02-1989	DE 3880468 A EP 0323997 A JP 2505268 B JP 2500330 T US 5591639 A US 5658759 A	27-05-1993 19-07-1989 05-06-1996 08-02-1990 07-01-1997 19-08-1997
WO 9720043	A	05-06-1997	AU 1024697 A EP 0874898 A	19-06-1997 04-11-1998
WO 9523868	A	08-09-1995	US 5639940 A AU 685250 B AU 1977095 A BR 9506944 A CA 2184004 A	17-06-1997 15-01-1998 18-09-1995 09-09-1997 08-09-1995

# INTERNATIONAL SEARCH REPORT

information on patent family members

International Application No

PCT/GB 98/02130

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9523868 A		CN 1144539 A	05-03-1997
		CZ 9602575 A	11-06-1997
		EP 0748386 A	18-12-1996
		FI 963419 A	02-09-1996
		HU 75552 A	28-05-1997
		JP 9509839 T	07-10-1997
		NO 963651 A	01-11-1996
		NZ 282601 A	24-06-1997
		PL 316115 A	23-12-1996
		SK 111896 A	09-07-1997
WO 8800239 A	14-01-1988	AT 159289 T	15-11-1997
		AU 605497 B	17-01-1991
		AU 7649087 A	29-01-1988
		DE 3752130 D	20-11-1997
		DE 3752130 T	09-04-1998
		DK 106588 A	29-02-1988
		EP 0274489 A	20-07-1988
		EP 0791652 A	27-08-1997
		JP 1500162 T	26-01-1989
		RU 2085587 C	27-07-1997
		US 5366894 A	22-11-1994
		US 5322775 A	21-06-1994
		US 5476995 A	19-12-1995
EP 0832981 A	01-04-1998	AU 618524 B	02-01-1992
		AU 1178488 A	18-08-1988
		AU 661696 B	03-08-1995
		AU 1941092 A	24-09-1992
		CA 1321764 A	31-08-1993
		EP 0279582 A	24-08-1988
		JP 63309192 A	16-12-1988
		US 5565362 A	15-10-1996
		US 5304489 A	19-04-1994



## INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

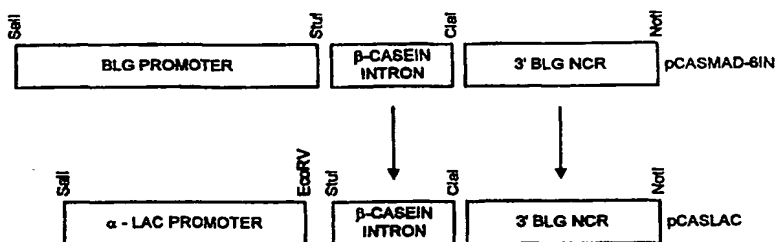
(51) International Patent Classification <sup>6</sup> : <b>C12N 15/00, 15/12, 15/85, 5/10, C07K 14/705, 14/78, 14/81, 16/00, A01K 67/027</b>		<b>A1</b>	(11) International Publication Number: <b>WO 99/03981</b>
			(43) International Publication Date: 28 January 1999 (28.01.99)
(21) International Application Number: PCT/GB98/02130		(81) Designated States: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, GH, GM, HR, HU, ID, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).	
(22) International Filing Date: 17 July 1998 (17.07.98)		<b>Published</b> <i>With international search report.</i> <i>Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.</i>	
(30) Priority Data: 9715064.3 17 July 1997 (17.07.97) GB			
(71) Applicant (for all designated States except US): PPL THERAPEUTICS (SCOTLAND) LIMITED [GB/GB]; Roslin, Edinburgh EH25 9PP (GB).			
(72) Inventors; and (75) Inventors/Applicants (for US only): COLMAN, Alan [GB/GB]; PPL Therapeutics (Scotland) Limited, Roslin, Edinburgh EH25 9PP (GB). GARNER, Ian [GB/GB]; PPL Therapeutics (Scotland) Limited, Roslin, Edinburgh EH25 9PP (GB). DALRYMPLE, Michael, Alexander [GB/GB]; PPL Therapeutics (Scotland) Limited, Roslin, Edinburgh EH25 9PP (GB).			
(74) Agents: CORNISH, K., V., J. et al.; Kilburn & Strode, 20 Red Lion Street, London WC1R 4PJ (GB).			

(54) Title: NUCLEIC ACID EXPRESSION CONSTRUCTS INCORPORATING A HETEROLOGOUS INTRON WHOSE NATURAL POSITION IS WITHIN THE 5'UNTRANSLATED REGION OF ITS GENE

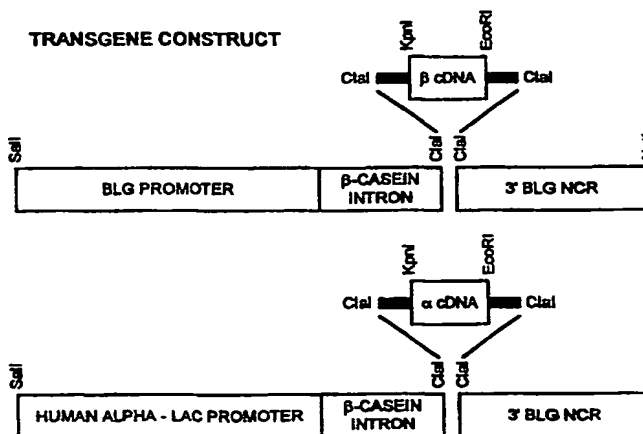
## (57) Abstract

A nucleic acid expression construct comprising: (a) a promoter; (b) an intron whose natural position is within the 5'-untranslated region of a gene from which it is derived; (c) a coding sequence; and (d) a 3'-flanking sequence wherein the intron (b) is not derived from the same gene as that from which either the promoter (a) or the protein-coding sequence (c) is derived and processes, vectors, hosts and uses involving such a construct to obtain *inter alia* an increase in the level of expression of the coding sequence.

## CLONING OF CASLAC



## TRANSGENE CONSTRUCT



**FOR THE PURPOSES OF INFORMATION ONLY**

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon		Republic of Korea	PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		